A NONLINEAR STOCHASTIC DYNAMIC SYSTEMS APPROACH FOR PERSONALIZED PROGNOSTICS OF CARDIORESPIRATORY DISORDERS

By

TRUNG QUOC LE

Bachelor of Engineering
Vietnam National University-Ho Chi Minh City
University of Technology
Ho Chi Minh City, Vietnam
2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in particular fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2013

A NONLINEAR STOCHASTIC DYNAMIC SYSTEMS APPROACH FOR PERSONALIZED PROGNOSTICS OF CARDIORESPIRATORY DISORDERS

Dissertation Approved:

Dr. Satish T. S. Bukkapatnam
Dissertation Adviser

Dr. William J. Kolarik

Dr. Zhenyu (James) Kong

Dr. Martin Hagan

ACKNOWLEDGMENTS

I would like to show my sincere gratitude to my academic advisor, Dr. Satish T.S. Bukkapatnam, for his constant encouragement, motivation, guidance, and financial support for my Ph.D study. I also would like to show my deepest appreciation to my co-advisor Dr. Ranga Komanduri who passed away. This dissertation is dedicated to my fond memory of him.

I also want to thank committee member Dr. Kolarik for introducing me to the real-time prognostics and risk analysis concepts. I would like to express my gratitude to Dr. Kong for teaching me reliability analysis and to Dr. Hagan for teaching me system identification, neural networks, and estimation theory. My appreciation goes to Dr. Bruce Benjamin and Dr. Brek Wilkins for introducing me to the physiology of the cardiovascular system.

Big thanks goes to all the friends and colleagues for their support and encouragement.

I would like to thank the National Science Foundation under Grant CMMI-0729552 and Grant CMMI-0830023, the Vietnam Education Foundation, and the AT&T Professorship for their financial support.

Heartfelt appreciation is owed to my wife Anh Tran, and my son Minh Le for their continuous support, motivation, and sympathy during my Ph.D.

Finally, I would like to dedicate this dissertation to my dear parents for their sacrifices, encouragement, and affection.

Name: TRUNG QUOC LE

Date of Degree: JULY, 2013

Title of Study: A NONLINEAR STOCHASTIC DYNAMIC SYSTEMS APPROACH FOR PERSONALIZED PROGNOSTICS OF CARDIORESPIRATORY DISORDERS

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract: This research investigates an approach rooted in nonlinear stochastic dynamic systems principles for personalized prognostics of cardiorespiratory disorders in the emerging point-of-care (POC) treatment contexts. Such an approach necessitates new methods for (a) quantitative and personalized modeling of underlying cardiovascular system dynamics to serve as a virtual instrument to derive surrogate (hemodynamic) signals, (b) high-specificity diagnostics to identify and localize disorders, (c) real-time prediction to provide forecasts of impending disorder episodes, and (d) personalized prognosis of the short-term variations of the risk, necessary for effective treatment decisions, based on estimating the distribution of the times remaining till the onset of an anomaly episode. The specific contributions of the dissertation work are as follows:

1. **Quantitative modeling for real-time synthesis of hemodynamic signals.** Features extracted from ECG signals were used to construct atrioventricular excitation inputs to a nonlinear deterministic lumped parameter model of cardiovascular system dynamics. The model-derived hemodynamic signals, personalized to an individual's physiological and anatomical conditions, would lead to cost-effective virtual medical instruments necessary for personalized POC prognostics.

2. **Random graph representation of the complex cardiac dynamics for disorder diagnostics.** The quantifiers of a random walk on a network reconstructed from vectorcardiogram (VCG) were investigated for the detection and localization of cardiovascular disorders. Extensive tests with signals from PTB database of PhysioNet databank suggest that locations of myocardial infarction can be determined accurately (sensitivity of ~88% and specificity of ~92%) from tracking certain consistently estimated invariants of this random walk representation.

3. **Nonparametric prediction modeling of disorder episodes.** A Dirichlet process based mixture Gaussian process was utilized to track and forecast the evolution of the complex nonlinear and nonstationary cardiorespiratory dynamics underlying of the measured signal features and health states. Extensive sleep tests suggest that the method can predict an impending sleep apnea episode to accuracies (R^2) of 83% and 77% for 1 step and 3 step-ahead predictions, respectively.

4. **Color-coded random graph representation of the state space for personalized prognostic modeling.** The prognostic model used the stochastic evolution of the transition pathways from a normal state to an anomalous state in the color-coded state space network to estimate the distribution of the remaining useful life. The prognostic model was validated using the data from ECG Apnea Database (Physionet.org). The model can predict the estimated time till a disorder (apnea episode) onset to within 15% of the observed times 1-45 min ahead of their inception.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Research motivation

Cardiovascular disease (CVD) is a leading cause of mortality, accounting for 30% of deaths worldwide (see Figure 1-1) [1]. On the basis of heart disease and stroke statistics updated in 2012, each day more than 3,400 Americans experience a new or recurrent stroke and 2,400 die of CVDs. The direct and indirect costs of medical devices, physicians, professionals, hospital services, and prescribed medication for CVD diagnosis and treatment were estimated at around $324 billion in 2009 [2]. Healthcare costs escalate exponentially with delay in detection of cardiovascular disorders [3]. The development of affordable and accessible medical instrumentation that supports clinical points-of-care (POC) diagnoses is essential for promoting early detection, thereby reducing the costs of treating cardiovascular disease. Furthermore, innovative research on prognostic approaches capable of forecasting early states of the diseases to facilitate effective preventive treatment is crucial for improving the patients' quality of life and alleviating socio-economic imbalances in the nation.

One of the most common goals in healthcare for the next decade is the transformation from reactive damage control to proactive and personalized wellness [4]. A new P4 medicine —Prediction, Prevention, Personalization, and Participation— based on integrating the concepts of a systems approach to diseases, emerging technologies, and advanced analytical tools, provides a personal basis for healthcare delivery to maximize wellness for each individual rather than

treating diseases (see Figure 1-2) [5]. P4 system provides deep insights into disease mechanisms, stratification of complex diseases, personalized treatments, and metrics for assessing wellness [4, 6]. It has been shown in the last 5 years that P4 system has significantly improved the way cancer is diagnosed and treated. Cardiovascular diseases and neurodegenerative disorders are the next targets for P4 system [6]. The realization of the P4 system in healthcare enables early diagnosis and real-time prognosis, especially for at-risk (e.g., critical care) populations, which can significantly lower treatment cost and reduce mortality and morbidity risks [3].

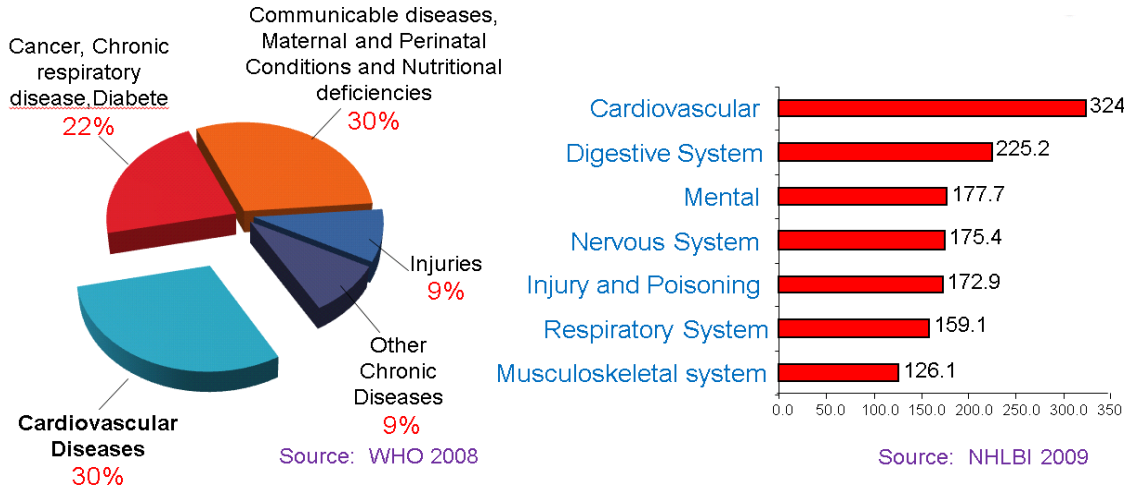


Figure 1-1 Major causes of death, and diagnosis and treatment costs of major diseases

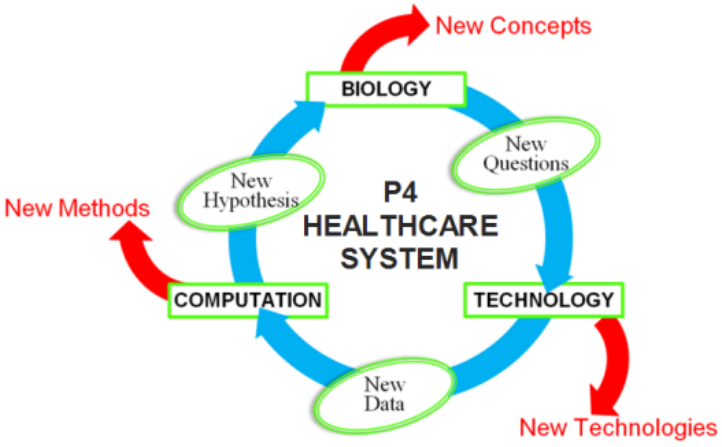


Figure 1-2 Cross-interdisciplinary fields constituting the P4 healthcare system

A major challenge to achieving P4 system in healthcare originates from the inadequacy of effective mathematical and computational platforms to capture the complexity of cardiovascular diseases [5]. Although efforts have been made in the last five years to develop a comprehensive approach to responding to these challenges, many difficulties have impeded the wide implementation of P4. One major limitation is the insufficiency of comprehensive concepts and practical schemes, including modeling approaches that can capture complex individualized physiological systems and disease mechanisms, methods for diagnosing distinct subtypes of diseases for an impedance match against proper drugs, and prediction and prognostic frameworks to provide reliable metrics for assessing wellness.

This research is an effort to address the technical barriers to a P4 system for cardiovascular diseases. Specifically, this research provides mathematical and computational schemes to address the *diagnostic and prognostic issues* in realizing the proactive and personalized diagnosis and treatment of cardiovascular disorders. The following aspects must be considered in developing such effective schemes: (1) a quantitative model that can capture the underlying cardiorespiratory couplings and generate noninvasive surrogate hemodynamic signals, (2) high-specificity diagnostic methods to identify and localize disorders, (3) real-time prediction methods that can drive advanced prognostic and preventive therapies, and (4) prognostic approaches that provide accurate risk indicators and survival assessments of the disease's progression.

1.2 Research objectives

This research addresses various phases for the development of prognostic schemes that support the implementation of the P4 system for cardiorespiratory disorder treatments. These phases include modeling the complex interactions between different physiological processes and control mechanisms unique for each individual's cardiovascular system, developing diagnostic

methods to detect and localize the disorders given the complex dynamics of the underlying cardiovascular system, and providing prediction and prognostic method to derive the distribution of time to abnormal conditions and reliable risk indicators to facilitate intervention therapies and preventive treatments. Interdisciplinary methods based on combining principles from nonlinear stochastic dynamic systems, random graph theory, and sensor-based modeling are used to characterize the complex dynamic, capture the coupling effects, and monitor long-term health transitions and short-term degradations of the underlying cardiovascular system.

Modeling the cardiovascular system, local diagnosis of cardiovascular diseases, and prediction of the pathological transitions must be addressed prior to the development of prognostic schemes. The research methodology needs to focus on the system dynamics approach that characterizes the coupled stochastic nonlinear nonstationary dynamics of the underlying cardiovascular system. With these emphases, the broad research objectives are as follows:

- Develop an effective data-driven cardiovascular model to provide surrogate hemodynamic signals for the individualized diagnosis and prognosis of cardiovascular disorders.
- Characterize nonlinear stochastic spatiotemporal dynamic of cardiac vectors and quantify the recurrence patterns of Vectorcardiogram (VCG) for high-specificity diagnosis of cardiorespiratory disorders.
- Develop a real-time prediction model to forecast the evolutions of the underlying coupled nonlinear and nonstationary cardiorespiratory dynamics that provides risk assessment for cardiorespiratory disorders.
- Develop and implement a prognostic framework for a P4 system of healthcare using the developed diagnostic, and prediction models.

1.3 Major contributions

The proposed research will provide mathematical and computational frameworks to develop effective prognostic schemes for the implementation of P4 system healthcare for the diagnosis and treatment of cardiorespiratory disorders. Several case studies focusing on diagnosis and prognosis of cardiorespiratory related diseases, e.g., myocardial infarction and obstructive sleep apnea have been used to validate the proposed models. The specific contributions of the dissertation are divided into two groups—methodology and application:

A. Methodology contributions

(1) Development of an overall prognostic scheme for the implementation of the P4 system of healthcare.

(2) A new real-time ECG-driven cardiovascular system model capable of capturing the mechanical-electrical coupling of the heart chambers, valves, and pulmonary and systemic circulations to generate surrogate hemodynamic signals with the applications to POC diagnosis.

(3) A stochastic representation of the complex cardiac excitation and propagation dynamics as a random walk on a network reconstructed from VCG signals.

(4) A prognostic model based on topological transitions of state space vectors for the preventive treatments of cardiovascular disorders.

B. Application contributions

(1) A myocardial infarction (MI) detection and localization model using topological and dynamic quantifiers of aperiodic and recurrent local transitions of VCG trajectories which can accurately (at a sensitivity of ~88% and specificity of ~92%) identify five typical MI types and healthy individuals.

(2) Quantifiers of the coupled nonlinear and nonstationary cardiorespiratory dynamics underlying the measured physiological signals used as significant features for the prediction and prognosis of sleep apnea onset.

1.4 Organization of the dissertation

This chapter presents the research motivation, research objectives, research contributions, and the dissertation organization. The rest of the dissertation is organized as follows:

Chapter 2: Background: Brief descriptions of the cardiovascular system physiology, cardiovascular electrical activities, and electrocardiogram (ECG) and vectorcardiogram (VCG) monitoring systems are provided. Next, the analysis tools including nonlinear analysis and recurrent quantification analysis (RQA) used to capture the dynamics underlying cardiovascular system are introduced. Finally, graph theory and a network representation of the VCG octant transition are presented.

Chapter 3: Overall methodology: An overview of the research methodology is followed by a list of the individual tasks composing the overall methodology. The research methodology is grouped into four parts: modeling of the cardiovascular system, diagnosis of local cardiovascular disorders, prediction of the cardiovascular system's dynamic evolution, and prognostics approach for cardiovascular disorders.

Chapter 4: Modeling of the cardiovascular system: The approach, implementation, and clinical validation of the ECG-driven cardiovascular system model are presented. Application of the model towards virtual instrumentation is also covered in this chapter.

Chapter 5: Diagnostics of local cardiovascular disorders: A probabilistic approach to quantifying the aperiodic patterns and the stochastic transitions of the cardiac vector in the 3-D octant state space is introduced. High-specificity diagnostic methods for identifying and localizing disorders using the VCG octant transition network are presented. A case study in

detecting and localizing 5 types of myocardial infarctions (MIs) and healthy individuals to validate the approach is also detailed.

Chapter 6: Prediction of the cardiovascular system's dynamic evolution: A feature extraction method and Dirichlet process based mixture Gaussian process (DPMG) prediction model capable of tracking and forecasting the evolutions of the cardiorespiratory dynamics captured from the measured physiological signals are presented. Data from obstructive sleep apnea (OSA) patients and healthy individuals collected from Physionet.org and from a wireless multisensory platform are used for model validation.

Chapter 7: Prognostics approach for cardiovascular disorders: The prognostic schemes necessary for the implementation of a P4 system for cardiovascular disease treatment are introduced in this chapter. The methodology, implementation, and validation of the model are investigated and validated through the case study of deriving the distribution of time to next sleep apnea onset with the data from the Apnea ECG database—Physionet.org.

Chapter 8: Conclusions and future work: This chapter summarizes the research contributions and the future work.

REFERENCES

[1] *Cardiovascular Diseases* Available: http://www.who.int/mediacentre/factsheets/fs317/en/index. html, WHO (2008).

[2] *Morbility and Mortality: 2009 Chart Book on Cardiovascular, Lung, and Blood Diseases*. Available: http://www.nhlbi.nih.gov/resources/docs/2009ChartBook.pdf, NHLBI (2009).

[3] R. Snyderman and Z. Yoediono, "Prospective Care: A Personalized, Preventative Approach to Medicine.," *Pharmacogenomics,* vol. 7, 5-9, Feb 2006.

[4] L. Hood, *et al.*, "Systems Biology and New Technologies Enable Predictive and Preventative Medicine," *Science,* vol. 306, 640-643, 2004.

[5] L. Hood and M. Flores, "A Personal View on Systems Medicine and the Emergence of Proactive P4 Medicine: Predictive, Preventive, Personalized, and Participatory," *New Biotechnology,* vol. 29, 613-624, 2012.

[6] Q. Tian, *et al.*, "Systems Cancer Medicine: Towards Realization of Predictive, Preventive, Personalized, and Participatory (P4) Medicine," *Journal of Internal Medicine,* vol. 271, 111-121, 2012.

CHAPTER 2

BACKGROUND

2.1 Physiology of the human cardiovascular system

The human cardiovascular system consists of three main components as shown in Figure 2-1: the heart, systemic circulation, and pulmonary circulation. The right side of the heart pumps blood through the lungs via the pulmonary circulation, and the left side of the heart pumps blood through the peripheral organs. The pumping action is enabled using a pulsatile two-chamber pump composed of an atrium and a ventricle. Each atrium acts as a weak primer pump for the ventricle, helping to move blood into the ventricle. The ventricles then supply the main pumping force that propels the blood either through the pulmonary circulation (the right ventricle) or through the peripheral circulation (the left ventricle). Complex physiological mechanisms of the heart controlled by the central nervous system cause recurring heart contraction signals called cardiac rhythms that transmit action potentials throughout the heart muscle to cause the heart's rhythmical beat.

The heart has two special functions: (1) generating rhythmical electrical impulses to cause rhythmical contraction of the heart muscle and (2) conducting these impulses rapidly through the heart. When this system functions normally, the atria contract about one sixth of a second ahead of ventricular contraction. Atrial contraction allows the filling of the ventricles before the blood is pumped through the lungs and peripheral circulation. These rhythmical pumping and conduction actions of the heart are affected by heart diseases such as myocardial infarction. Myocardial

infarction often results from damage to the heart muscle cells (myocytes) due to the interruption of the blood supply. It is often manifested as a bizarre heart rhythm or an abnormal sequence of contractions of the heart chambers, severely affecting the pumping effectiveness of the heart, even to the extent of causing death.

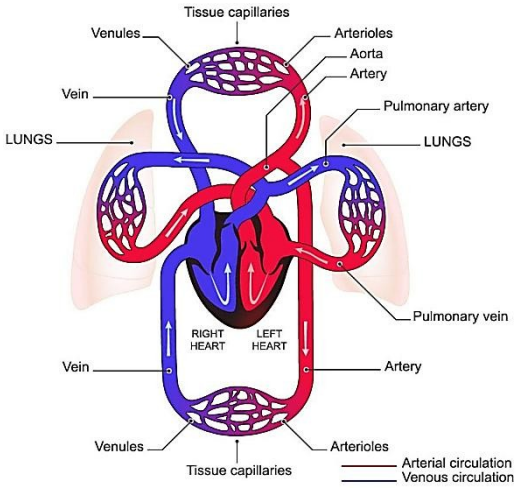


Figure 2-1 Structure of the cardiovascular system [1]

Systems have been designed to monitor cardiac activities towards timely detection of CVD. Most of the systems are based on measuring the potential electrical changes when the cardiac impulse passes through the heart. If electrodes are placed on the skin on opposite sides of the heart, electrical potentials generated by the current can be recorded: such a recording is known as an electrocardiogram (ECG). The ECG system developed by Augustus Waller in 1889 and improved by Willem Einthoven in 1901 is still in use and serves as the gold standard for clinical diagnosis of cardiovascular disorders. In 1904, Einthoven developed "Einthoven triangle," which measures the three channels of ECG signals (Leads I-III) and derives the direction of the electric heart vector.

A normal ECG waveform is composed of a P wave, a QRS complex, and a T wave (see Figure 2-2). The QRS complex is often, but not always, three separate waves: the Q wave, the R

wave, and the S wave. The P wave is caused by electrical potentials generated when the atria depolarize before atrial contraction begins. The QRS complex is caused by potentials generated when the ventricles depolarize before contraction, that is, as the depolarization wave spreads through the ventricles. Therefore, both the P wave and the components of the QRS complex are depolarization waves. The T wave is caused by potentials generated as the ventricles recover from the state of depolarization and is known as a repolarization wave. This process normally occurs in the ventricular muscle 0.25 to 0.35 second after depolarization. Thus, the electrocardiogram is composed of both depolarization and repolarization waves.

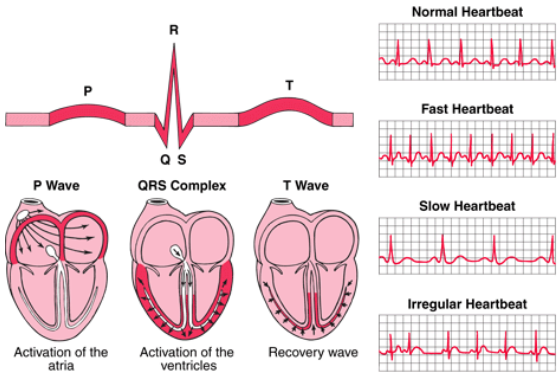


Figure 2-2 Typical ECG signal, mechanical event diagrams, and various ECG signal waveforms corresponding to heartbeat conditions

In 1956 Ernest Frank redesigned the lead configuration by proposing three pairs of leads to measure the electric heart vector in the Cartesian coordinate system. This redesign led to the advent of the vectorcardiogram (VCG). The VCG signals capture the electrical potential of the heart as an electric heart vector in a three orthogonal coordinate system as shown in Figure 2-3. Dower et. al. [2, 3] showed that the linear transformation between a VCG and 12-lead ECG preserves useful information regarding the heart dynamics. In addition, Edenbrandt and Pahlm [4] proved that VCG criteria for the diagnosis of, for example, myocardial infarction and right ventricular hypertrophy, are superior to the corresponding 12-lead ECG.

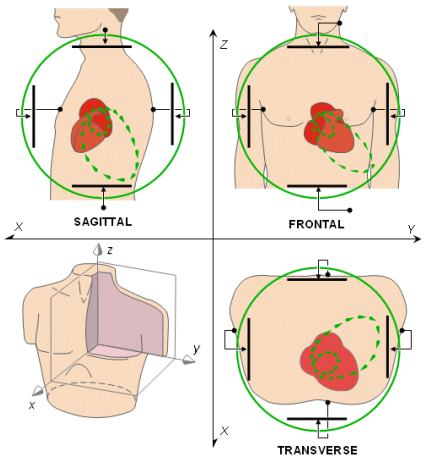


Figure 2-3 Electrode placements in a VCG measurement system

2.2 Primer on nonlinear dynamic analysis

Many natural phenomena are subject to nonlinear dynamics. The study of nonlinear dynamics is typically restricted to systems which respond disproportionately (nonlinearly) to initial conditions or perturbing stimuli. Nonlinear systems may exhibit chaos, which is classically characterized as sensitive dependence on initial conditions. A typical nonlinear dynamic system can be described as follows:

$$d\underline{x} = \underline{F}(\underline{x})dt + \underline{g}(\underline{x})d\underline{\beta}\ ,$$

where $\underline{x}(t)$ is a $d_{\underline{x}}$-dimentional state space vector, $\underline{F}(\cdot)$ is a nonlinear vector field, t is the time, and the $\underline{g}(\underline{x})d\underline{\beta}$ term accounts for the dynamic noise of extraneous phenomena [5, 6]. The hypothesis that cardiac rhythms are associated with chaotic dynamics has motivated the investigation of continuous ECG and VCG signals using nonlinear dynamic analysis. The physiological cardiovascular regulation is known to be associated with the parasympathetic and sympathetic control of cardiac dynamics. Thus, the nonlinear measures of ECG and VCG signals recorded during different cardiac cycles can capture the behavior of the complex cardiovascular system.

Phase space (or state space) methods provide powerful tools for the analysis of the dynamics of a nonlinear system. The phase space is reconstructed from the delayed coordinates of the measurements $y(t)$ and is given as:

$$\underline{x}(t_i) = [y(t_i), y(t_i + \tau), y(t_i + 2\tau), \dots, y(t_i + (m-1)\tau)],$$

where m is the embedding dimension and τ is the delay time. The minimum sufficient embedding dimension m is defined by the false nearest neighbor method [7, 8] and the optimal τ is selected by minimizing the mutual information function [9].

Recurrence is the fundamental characteristic of state vectors that can be exploited to capture the dynamics of the underlying system in the phase space. A powerful tool for visualizing and characterizing recurrence properties is the recurrence plot [10-12]. The recurrence plot is formed by calculating the distance from each state vector to all other state vectors and mapping the distance to a color scale (continuous or binary). The recurrence plot is expressed as:

$$R_{i,j}(\varepsilon) = \Theta\left(\varepsilon - \left\|\vec{x_1} - \vec{x_j}\right\|\right);\ i, j = 1, \dots, N,$$

where N is the number of measured points $\vec{x_1}$, ε is a threshold distance, $\Theta(\cdot)$ is the Heaviside function, and $\|\cdot\|$ is a norm.

Recurrence quantification analysis (RQA) of the recurrence plot has been widely used to capture the aperiodic and recurrent features of the VCG trajectory and the laminar or chaotic behaviors of heart rate variability in the phase space. The RQA defines measures based on the distribution of recurrence point density, diagonal structures, and vertical structures in the recurrence plot. The definitions of various RQA measures and their relationships with cardiac system dynamics are summarized in Table 2-1.

Table 2-1 Mathematical definitions of the RQA quantifiers and their relationships to cardiac system dynamics

RQA Features	Description
Recurrence Rate	$\mathrm{RR} = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}$ $R_{i,j}$: the indicator of whether the point j is a recurrent point of i (i.e., lies in an ε-neighborhood of i) in the state space. N: the length of the time series. Recurrence rate characterizes the global aperiodicity of cardiovascular activities and it is closely related to the heart rate dynamics [12]
Determinism	$\mathrm{DET} = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}}$ $P(l)$: probability distribution (estimated from histogram transformations) of the lengths l of the diagonal lines. Determinism measures the repeating or deterministic patterns in the heart dynamics and shows how well the circulatory system functions [10, 11].
Average diagonal length	$\mathrm{DIA} = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)}$
Longest diagonal length	$\mathrm{LMAX} = \max(\{l_i; i = 1, \dots, N_l\})$ N_l: the total count of diagonal lines in the recurrence plot. LMAX indicates stability of heart dynamics and small LMAX implies two close cardiac vectors in state space will diverge quickly from each other.
Entropy	$\mathrm{ENTR} = -\sum_{l=l_{min}}^{N} P(l)\ln P(l)$ The predictivity of heart activity decreases with increasing entropy
Laminarity	$\mathrm{LAM} = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}$ $P(v)$: probability distribution of the length v of the vertical lines. LAM values is proportional to the time the heart takes to move from one activity to another and it provides non-stationary information for the heart system [10].
Trapping time	$\mathrm{TT} = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)}$ TT measures how long the cardiac vector remains in a specific state.
Longest vertical line length	$\mathrm{LVM} = max(\{v_i; i = 1, \dots, N_v\})$ N_v: the total count of vertical lines in the recurrence plot. LMAX is used to detect and quantify the laminar phases (chaos-chaos transitions) before a life-threatening cardiac arrhythmia occurs [10, 13].
Recurrent time type 1	$\mathrm{RT1} = \frac{1}{N}\sum_{i=1}^{N} T_i^{(1)}$ $T_i^{(1)}$: the average of the minimum time difference between points in the neighborhood of a point i on the reconstructed trajectory [14].
Recurrent time type 2	$\mathrm{RT2} = \frac{1}{N}\sum_{i=1}^{N} T_i^{(2)}$ $T_i^{(2)}$: the average return time (i.e., the minimum time difference between the recurrence points in the neighborhood of point i on the reconstructed trajectory with all successive time points excluded) [14].
Recurrence period entropy density	$\mathrm{RENT} = \lim_{\varepsilon\to\infty}\lim_{m\to\infty}\frac{1}{\tau}\ln\frac{C^m(\varepsilon)}{C^{m+1}(\varepsilon)}$ Kolmogorov entropy estimated from the recurrence plot [15]. RENT is used to quantify deterministic structure of the system.
Transitivity	$\mathrm{TRAN} = \frac{\sum_{k=1}^{N} W_k}{\sum_{k=1}^{N} \Delta_k}$ W_k: the number of triangle links. Δ_k: the number of connected triplets of links in the network with the phase space vectors as the nodes and the recurrences as the links. TRAN measures the psychophysiological variables of heart rate variability with circadian rhythmicity [16]

2.3 A graph-theory perspective of nonlinear dynamic systems

Graph theory offers a new way to quantitatively characterize the topological and dynamic properties of a complex system. A graph $G(V, E)$ of a network consists of a set of vertices (or nodes) V and a set of edges (or connections) E. The presence of the edges between two vertices represents any type of interaction or connection between the vertices. The interpretation of the edges depends on the types of connections modeled e.g., correlations, coherence, and mutual information. The information from the graph connectivity is completely described by an adjacency matrix A. Each entry A_{ij} stands for the existing edge between vertices i and j; i.e., if there is a connection between vertices i and j, $A_{ij} = 1$; otherwise $A_{ij} = 0$. Graphs can be categorized into different types: undirected, when information can flow in both directions along edges connecting vertices or directed, when information can flow in only one direction. Graphs can be unweighted, when the edges have the same significance or weighted, if weights are assigned to each edge. Figure 2-4 shows an example of a network representation of the functional connectivity of the human brain.

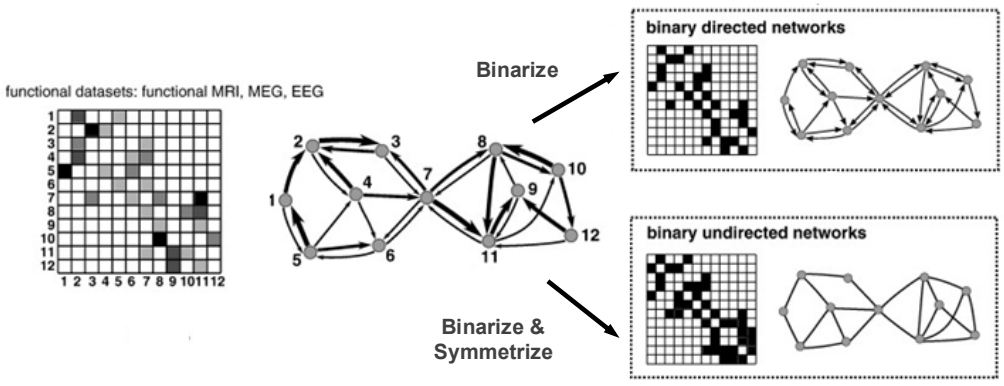


Figure 2-4 Construction of the brain network from a functional connectivity data set extracted from neuroimaging (fMRI) or neurophysiological (MEG, EEG) signals.

The network measures characterize global and local connectivity of the complex network. Definitions of various network measures and their interpretations are summarized as follows:

Table 2-2 Description of basic network measures

Basic concepts and measures	
Basic concepts and notation	N: is the set of all nodes in the network, and n is the number of nodes L: is the set of all links in the network, and l is number of links (i,j): is a link between nodes i and j $(i,j \in N)$ a_{ij}: is the connection status between i and j w_{ij}: is the weight of the link (i,j)
Degree	Degree of a node i $$d_i = \sum_{j \in N} a_{ij}$$
Shortest path length	Shortest path length (distance), between nodes i and j $$d_{ij} = \sum_{a_{uv} \in g_{ij}} a_{uv}$$ where g_{ij} is the shortest path (geodesic) between i and j. Note that d_{ij}=∞ for all disconnected pairs i, j.
Number of triangles	Number of triangles around a node i $$t_i = \frac{1}{2} \sum_{j,h \in N} a_{ij} a_{ih} a_{jh}$$
Measures of integration	
Characteristic path length	Characteristic path length of the network [17] $$L = \frac{1}{n} \sum_{i \in N} L_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}}{n-1}$$ where Li is the average distance between node i and all other nodes.
Global efficiency	Global efficiency of the network [18] $$E = \frac{1}{n} \sum_{i \in N} E_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^{-1}}{n-1}$$ where E_i is the efficiency of node i.
Clustering coefficient	Clustering coefficient of the network [17] $$C = \frac{1}{n} \sum_{i \in N} C_i = \frac{1}{n} \sum_{i \in N} \frac{2t_i}{k_i(k_i-1)}$$ where C_i is the clustering coefficient of node i (C_i= 0 for k_i < 2).
Transitivity	Transitivity of the network [19] $$T = \frac{\sum_{j,h \in N} 2t_i}{\sum_{i \in N} k_i(k_i-1)}$$ Transitivity is not defined for individual nodes.
Measures of segregation	
Local efficiency	Local efficiency of the network [18] $$E_{loc} = \frac{1}{n} \sum_{i \in N} E_{loc,i} = \frac{1}{n} \sum_{i \in N} \sum_{j,h \in N, i \neq j} \frac{a_{ij} a_{ih} \left[d_{jh}(N_i)\right]^{-1}}{k_i(k_i-1)}$$ where $E_{loc,i}$ is the local efficiency of node i, and $d_{jh}(N_i)$ is the length of the shortest path between j and h, that contains only neighbors of i.
Modularity	Modularity of the network [20] $$Q = \sum_{u \in M} \left[e_{uu} - \left(\sum_{v \in M} e_{uv} \right)^2 \right]$$ where the network is fully subdivided into a set of non-overlapping modules M, and e_{uv} is the proportion of all links that connect nodes in module u with nodes in module v.
Measures of centrality and resilience	
Closeness centrality	Closeness centrality of node i [21] $$L_i^{-1} = \frac{n-1}{\sum_{j \in N, j \neq i} d_{ij}}$$

A random walk has been used to analyze the topological properties and dynamic features of graphs representing complex systems such as the World Wide Web, social networks, food webs, and interacting biological networks [13, 22-24]. Given a graph and starting point, a random walk on a graph is defined as a sequence of nodes whose neighbors are selected randomly. It is a time-reversible finite Markov chain [22]. Let $G(V, E)$ be a connected graph with starting node v_0. If at the t^{th} step, we are at node v_t, the sequence of random nodes $\{v_t: t = 0,1, \ldots.\}$ is a Markov chain. The probability matrix of this Markov chain is $M = (p_{ij})_{i,j\in E}$ where p_{ij} is defined as:

$$p_{ij} = \begin{cases} \frac{1}{d(i)}, & if\ i,j \in E \\ 0, & otherwise \end{cases},$$

where $1/d(i)$ is the transition probability from node i to node j. Let A be the adjacency matrix of G and D the diagonal matrix with $(D)_{ii} = 1/d(i)$; then $M = DA$. The random walk rule can be expressed as:

$$P_{t+1} = M^T P_t = (M^T)^t P_0,$$

where $P_t(i) = Prob(v_t = i)$. It is noted that p_{ij}^t is the probability that starting at i, we reach j in t steps and p_{ij}^t is equal to the ij entry of M^t.

Three important measures for the quantitative theory of random walks are access time (hitting time), cover time, and mixing rate:

(a) Access time, H_{ij}, is the expected number of steps before node j is visited, starting from node i. The sum $K(i,j) = H(i,j) + H(j,i)$ is called the commute time, which is the expected number of random walk steps starting at i and visiting node j before reaching node i again.

(b) Cover time, C, is the expected number of steps to reach all of the nodes in the graph given a starting distribution. If the starting distribution is not specified, cover time is the maximum values of the cover time from every node in the graph.

(c) Mixing rate, μ, is the expected number of steps required for v_t to converge to a stationary distribution. It quantifies how fast the random walk converges to its limiting distribution.

REFERENCES

[1] A. C. Guyton, *Textbook of Medical Physiology*, 8th ed.: Philadelphia: Saunders, 1991.

[2] G. E. Dower, "A Lead Synthesizer for the Frank System to Simulate the Standard 12-Lead Electrocardiogram," *Journal of Electrocardiology,* vol. 1*,* 101-116, 1968.

[3] G. Dower, "The Ecgd: A Derivation of the Ecg from Vcg Leads," *Journal of Electrocardiology,* vol. 17*,* 189-191, 1984.

[4] L. Edenbrandt and O. Pahlm, "Vectorcardiogram Synthesized from a 12-Lead Ecg: Superiority of the Inverse Dower Matrix," *Journal of Electrocardiology,* vol. 21*,* 361-367, 1988.

[5] S. T. Bukkapatnam*, et al.*, "Dependence of Computed Trajectory on Step-Size in a Nonlinear Dynamic System: An Investigation into Cutting Tool Dynamics," *IIE transactions,* vol. 27*,* 519-529, 1995.

[6] S. T. Bukkapatnam*, et al.*, "The Neighborhood Method and Its Coupling with the Wavelet Method for Signal Separation of Chaotic Signals," *Signal Processing,* vol. 82*,* 1351-1374, 2002.

[7] F. Takens, "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, ed: Springer, 1981, pp. 366-381.

[8] M. B. Kennel, *et al.*, "Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction," *Physical review A,* vol. 45, 3403, 1992.

[9] A. M. Fraser and H. L. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information," *Physical Review A,* vol. 33, 1134, 1986.

[10] N. Marwan, *et al.*, "Recurrence-Plot-Based Measures of Complexity and Their Application to Heart-Rate-Variability Data," *Physical Review E,* vol. 66, 026702, 2002.

[11] J. P. Zbilut, *et al.*, "Recurrence Quantification Analysis as a Tool for Nonlinear Exploration of Nonstationary Cardiac Signals," *Medical Engineering & Physics,* vol. 24, 53-60, 2002.

[12] H. Yang, "Multiscale Recurrence Quantification Analysis of Spatial Cardiac Vectorcardiogram Signals," *Biomedical Engineering, IEEE Transactions on,* vol. 58, 339-347, 2011.

[13] T. Le, *et al.*, "Topology and Random Walk Network Representation of Cardiac Dynamics for Localization of Myocardial Infarction," *IEEE Transactions on Biomedical Engineering,* 2013.

[14] J. Gao, "Recurrence Time Statistics for Chaotic Systems and Their Applications," *Physical Review Letters,* vol. 83, 3178-3181, 1999.

[15] P. Faure and H. Korn, "A New Method to Estimate the Kolmogorov Entropy from Recurrence Plots: Its Application to Neuronal Signals," *Physica D: Nonlinear Phenomena,* vol. 122, 265-279, 1998.

[16] C. Heinze, *et al.*, "The Relation of Heart Rate Variability Measures to Psychophysiological Variables with Circadian Rhythmicity," *Biomed Tech,* vol. 57, 1, 2012.

[17] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World'networks," *Nature,* vol. 393, 440-442, 1998.

[18] V. Latora and M. Marchiori, "Efficient Behavior of Small-World Networks," *Physical Review Letters,* vol. 87*,* 198701, 2001.

[19] M. E. Newman, "The Structure and Function of Complex Networks," *SIAM review,* vol. 45*,* 167-256, 2003.

[20] M. E. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Physical Review E,* vol. 69*,* 026113, 2004.

[21] L. C. Freeman*, et al.*, "Centrality in Valued Graphs: A Measure of Betweenness Based on Network Flow," *Social networks,* vol. 13*,* 141-154, 1991.

[22] L. Lovasz, "Random Walks on Graphs: A Survey," *Combinatorics,* vol. 2*,* 1993.

[23] H. Tong*, et al.*, "Fast Random Walk with Restart and Its Applications," 2006.

[24] F. Fouss*, et al.*, "Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 19*,* 355-369, 2007.

CHAPTER 3

RESEARCH METHODOLOGY

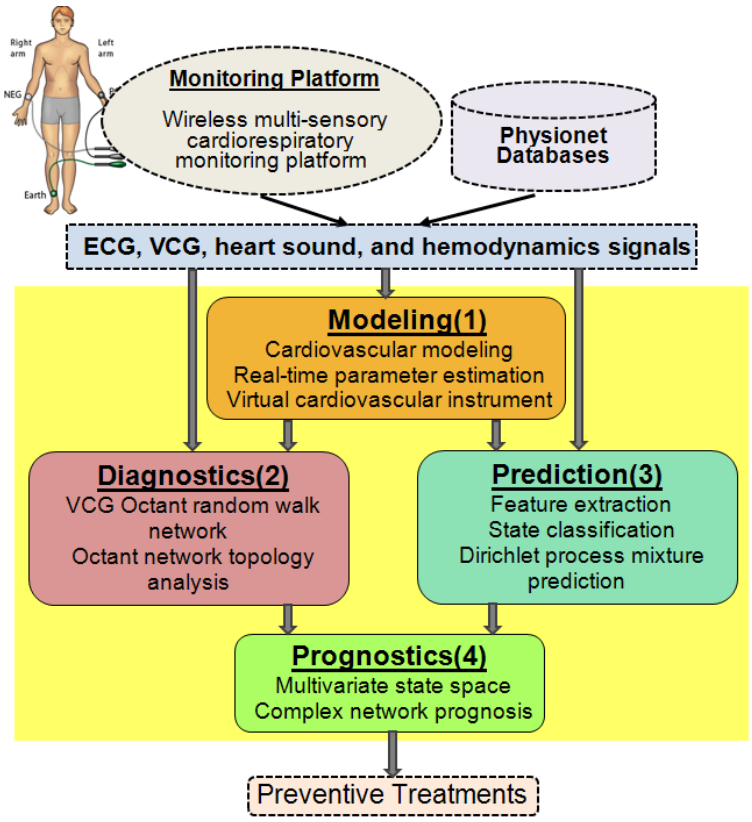


Figure 3-1 Overview of research methodology

This chapter outlines the road map of the research reported in subsequent chapters. Figure 3-1 portrays a schematic of the overall methodology employed to develop prognostic schemes for the implementation of the P4 system in cardiovascular disease diagnosis and treatment. Four modules constitute the overall research methodology as follows:

(i) Modeling of the cardiovascular system

(ii) Diagnosis of local cardiovascular disorders

(iii) Prediction of incipient disorder episodes

(iv) Prognostics approach for cardiovascular disorders

Physiological signals such as ECG, VCG, heart sounds, and hemodynamics collected from two data sources—Physionet Databases (physionet.org) and a wireless multisensory platform (COMMSENS Oklahoma State University lab)—have been used as the input information for two modeling, diagnosis modules. The lump parameter model of the cardiovascular system developed in the modeling module is used as a virtual cardiovascular instrument to generate other surrogate hemodynamic signals without the need for expensive instrumentation and/or invasive clinical procedures. The diagnosis module suggests a new, high-specificity diagnostic method to identify and localize cardiovascular disorders using the stochastic transition quantifiers of the cardiac vectors in the octant space. The prediction module provides a method for real-time tracking and forecasting of the evolutions of the underlying dynamics of the surrogate signals and local information of the disorder states generated from the model and diagnosis modules. The prognostic module, followed by the diagnosis and prediction of the disorder's states, provides a method for estimating the risk and provides the reliability assessment. The risk indicators from the prognostic model facilitate precise and timely preventive treatments and personalized therapies. The combination of these four modules constitutes a comprehensive prognostic scheme, which is necessary for the implementation of the P4 system.

3.1 Modeling the cardiovascular system

The first part of the research methodology involved developing a data-driven cardiovascular system model capable of generating multiple synchronized hemodynamic signals. A real-time

lumped parameter cardiovascular dynamics model with the atrioventricular activation function derived from ECG features was used to capture the physiological mechanisms and interactions in the cardiovascular system. The model represented the coupled dynamics of the heart chambers, valves, and pulmonary and systemic blood circulation loops in the form of nonlinear differential equations. The features extracted from ECG signals, including the time profile and respiratory components, were used to estimate the timings and amplitudes of the atrioventricular activation input functions.

To capture the unique characteristics of the cardiovascular system and real-time rendering of the hemodynamic signals from the measured ECG signal, an offline statistical model was used to map the model parameters to appropriate ECG features. A set of significant parameters, including the elastance characteristics, respiratory coupling, and gain and offset of model blood pressures, was selected for the parameter tuning. While the model-generated pressure waveforms can be compared with those from the actual recordings, only certain extreme values of the waveforms were considered clinically important. We have developed a method based on Anderson–Darling statistics and Kullback–Leibler divergence to compare the clinical measures (i.e., systolic and diastolic pressures) estimated from model waveform-extrema with those from actual measurements. Detailed descriptions of the model components, the parameter estimation approach, and the clinical validation procedures are presented in Chapter 4.

3.2 Diagnosis of local cardiovascular disorders

The second part of the research methodology involved developing a diagnostic method for detecting and localizing cardiovascular disorders. In this part, a probabilistic approach was used to quantify the aperiodic pattern and the stochastic transitions of the VCG trajectory in the 3-D octant state space. The variations of the transitions, which may be viewed as the output of the

cardiac process, were used to quantify the spatiotemporal dynamics underlying the cardiovascular disorders.

A high-specificity diagnostic method for identifying and localizing heart impairment is summarized in four steps. First, the transitions of the cardiac vector among eight octants (defined by embedding the VCG signals in the Euclidean three-dimensional coordinate system) were represented as a network with the node set as the octants and the edge set as the plausible transitions. Second, a random walk process was used to estimate the number of transitions (signal lengths) necessary for consistent estimation of the octant network quantifiers. Third, various local, residence, transitional, and topological features were extracted from the network with the signal lengths determined from previous steps for the localization of the disorder. Finally, hierarchical classification and regression tree (CART) models were used to classify different types of diseases. The development of the model and a case study for the detection and localization of six typical types of MIs (282 recordings) from the Physionet PTB Database are described in Chapter 5.

3.3 Prediction of incipient disorder episodes

The third part of the research methodology involved developing a prediction model capable of tracking and forecasting the evolutions of the cardiorespiratory dynamics captured from the measured physiological signals.

The prediction was performed through three phases—feature extraction, feature prediction, and disease classification. In the first phase, features were extracted on the basics of recurrence quantification analysis (RQA), which can capture the coupled nonlinear and nonstationary cardiorespiratory dynamics underlying the measured signals gathered from a custom-designed wireless wearable multisensory suite. In the second phase, a Dirichlet process based mixture

Gaussian process (DPMG) prediction model was employed to forecast the onset of the disorder episodes based on analyzing the complex evolutions of the extracted features. Finally, a Support Vector Machine (SVM) classification model was used to discern the normal and the abnormal states from the predicted values.

In order to facilitate the method implementation, we developed a prototype of a wireless multisensory platform capable of synchronously gathering multiple heterogeneous signals, including VCG, ECG, sound, and respiration, and wirelessly transmitting the data to a host computer for on-line prediction and subsequent therapeutic decision support. The prediction model was tested with two sources of data: (1) the Apnea-ECG database from Physionet.org and (2) a wireless multisensory platform developed at COMMSENS lab at Oklahoma State University. The prediction approach, the wireless multisensory platform, and the case study for the prediction of OSA onset are described in Chapter 6.

3.4 Prognostics approach for cardiovascular disorders

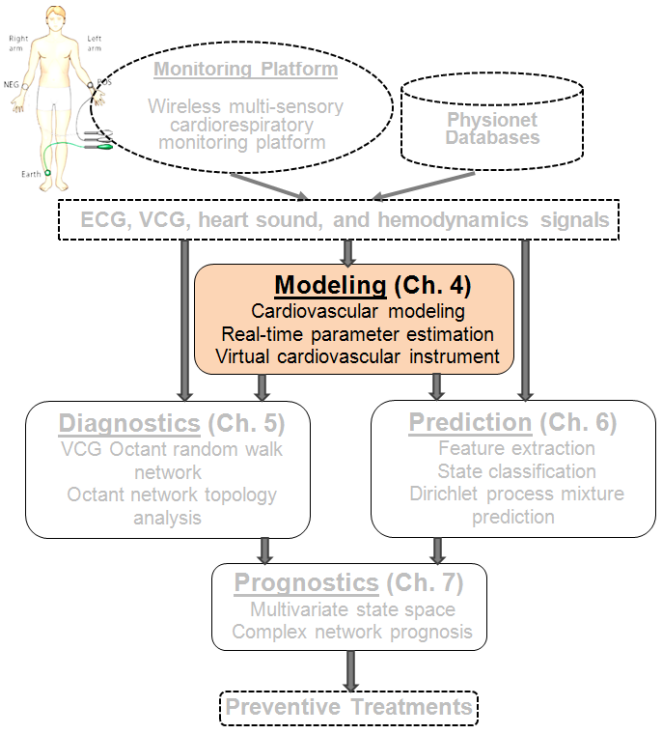
The final part of the research methodology combined three previous modules—modeling, diagnosis, and prognosis—to develop a prognostic scheme for the implementation of the P4 system for cardiovascular disease treatments.

The prognostic scheme derived the distribution of the time to failure of new observations collected from the heart rate variability (RR interval) signals. Two features—power spectrum density and longest vertical line of recurrence plot derived from RR interval signals using the sliding window concept—were used to reconstruct the multivariate state space. The embedded feature state space was partitioned into various clusters using a Dirichlet process. The state space was represented as a directed graph $G(V, E)$ where the node set V was the state vectors and the edge set E was the transition in the state space. An eigen projection method was employed to

account for the scattering and overcrowding of the adjacent nodes in the state space. The state vectors were projected to the 2nd and 3rd smallest Laplacian-eigen vectors that were subjected to the force directed strategy. Distribution of the time to abnormal onset of a new normal observation was estimated by considering the stochastic evolution of the normal state vectors to the abnormal state in the state space. For a case study, data from the Apnea-ECG database Physionet.org were investigated to validate the prognostic performance. The details prognostics approach are presented in Chapter 7.

CHAPTER 4

LUMPED PARAMETER NONLINEAR CARDIOVASCULAR SYSTEM MODELING FOR POC PERSONALIZED SIGNAL GENRATION



Numerical modeling of the cardiovascular system provides anatomical and physiological quantifications of the patient-specific cardiovascular system dynamics. Furthermore, cardiovascular models based on noninvasive medical signals could provide invaluable data on the in vivo environment where cardiovascular devices are expensive or inaccessible. We present an approach to deriving a real-time, lumped parameter cardiovascular dynamics model that uses

features extracted from online ECG to generate certain surrogate hemodynamic signals. The surrogate hemodynamic signals can be utilized as the input signals for the diagnostics, prediction, and prognostic models in chapters 5, 6, and 7, respectively. The model was tested using hemodynamic signals from the PhysioNet MGH/MF Waveform database. The results suggest that the model can capture the salient time and frequency patterns of the measured central venous pressure, pulmonary arterial pressure, and respiratory impedance signals (R^2>0.65). The results indicate the potential of a virtual instrument platform that can support more responsive, and cost-effective diagnostic medical equipment especially for point-of care diagnosis and personalized treatment.

4.1 Introduction

Cardiovascular diseases are the leading cause of mortality in the U.S. [1]. Delivery of cardiovascular healthcare, especially to rural and isolated communities, remains a major challenge despite over $10.83 billion spent annually on medical devices for cardiovascular diagnostics and treatment [2]. The development of affordable and accessible medical instrumentation is essential for promoting early diagnosis, thereby reducing cardiovascular disease treatment costs. A virtual cardiovascular instrument (VCI) in which the multiple data sets necessary for clinical diagnostics are generated through transformation of one measured signal can obviate the need for expensive clinical diagnostic instrument suites and thus has the potential to alleviate healthcare cost and access issues. Analytical lumped parameter [3-5], and/or computational (e.g., finite element) [6-8] modeling of the complex interactions among electrical, mechanical, and chemical processes that underpin cardiovascular dynamics is essential to the development of a VCI. Lumped parameter models offer lower computational cost and a straightforward physical interpretation of the dynamic interactions among the elements of a cardiovascular system. The present study extends our recent work [9] in developing a lumped

parameter model to relate the states of a cardiovascular system to online ECG signal features. Analogous to the virtual instruments used in engineering measurement practice [10], the model utilizes a single channel of the ECG to generate multiple virtual hemodynamic signals including pressure, volume, respiratory impedance, and blood flow rate.

Similar mathematical model-based platforms, such as LiDCOplus [11] and Picco2 [12] are used in critical care settings to provide estimates of blood pressure, stroke volume variations, and cardiac output. These platforms rely on invasive arterial pulse pressure measurements to estimate hemodynamic signals. They are, therefore, limited in their applications for easy bedside monitoring across a range of medical settings. The contributions of the present approach emerge from the methods to (a) derive atrioventricular activation functions based on relating the measured ECG signal events to the functions of various aspects of the cardiovascular processes; (b) estimate the parameters of a cardiovascular dynamics model so that real-time rendering of the hemodynamic signals from the measured ECG is facilitated; and (c) test the similarity of clinically relevant systolic and diastolic pressures extracted from the model to those from actual measurements. Although the approach is limited to capturing certain timing- and magnitude-related parameters of the activation functions from ECG signals and does not explicitly consider the effects of various control and regulation mechanisms of the circulatory system, our experimental investigations suggest that the model can capture the salient time and frequency patterns of certain measured hemodynamic signals (e.g., central venous pressure (CVP), pulmonary arterial pressure (PAP), and respiratory impedance(RI)). and provide real-time estimates of systolic and diastolic pressures—the key indicators in clinical practice. We anticipate that the present approach will spur further extensions that can lead to VCIs capable of generating certain hemodynamic signals relevant for diagnostics without the need for expensive instrumentation and/or invasive clinical procedures. This paper is organized as follows: Sec. 4.2

summarizes previous research on cardiovascular system modeling, Sec. 4.3 presents the formulation of the cardiovascular dynamics model and the parameter estimation procedures, Sec. 4.4 describes the implementation of the heart model and comparisons of model outputs with measured hemodynamic signals, and Sec. 4.5 contains a discussion of the model applicability and concluding remarks.

4.2 Background and literature review

The advent of digital computers has allowed the quantitative modeling of the complex mechanisms and interactions in the cardiovascular system [13, 14]. Notably, Avolio [15] employed impedance characteristics of arterial segments in a lumped parameter fluid dynamics model of the circulation system. Rideout and Dick's [16] difference-differential equation circulation system model was based on segmenting the arterial flow along the axial and radial directions of a cylindrical coordinate system. Guyton *et al.*'s seminal model [17] used more than 350 compartments to analyze the underlying interactions and predict the circulation states of the cardiovascular system. Subsequently, research efforts to simulate cardiovascular pathologies based on Guyton's model have been reported [18-20]. In many of these lumped parameter models, atrioventricular activation functions can be characterized effectively in terms of a time-varying elastance $e(t)$ [21, 22] that describes the average instantaneous variation of pressure (P) for a unit change in the volume (V) of an atrium (a) or ventricle (v), i.e., $e_{\mathrm{a,v}}(t) = \langle \frac{\partial P}{\partial V} \rangle_{\mathrm{a,v}}$. Normalized $e(t)$ of the left ventricle has been shown [23] to be fairly independent of loading conditions, contractile state, and the heart rate. Subsequently, Burkhoff *et al.* [24] and Kass *et al.* [25] showed that P-V characteristics are nonlinear during the *ejection phase* of a cardiac cycle, when intraventricular pressures exceed the aortal and pulmonary arterial pressures to propel blood out of the ventricles, as well as in the *isovolumic relaxation phase*, when the

intraventricular pressures decrease rapidly to cause the aortic and pulmonary valves to abruptly close. Senzaki *et al.* [26] presented a method to estimate the end-systolic elastance of the left ventricle from a single cardiac cycle. Klotz *et al.* [27] used Doppler-echocardiography and radio-nuclide ventriculography to capture *P-V* characteristics at every heart-beat. These earlier methods to estimate $e(t)$ used data collected from expensive instrumentation under well-defined conditions from *ex vivo* or *in vivo* experiments on an animal (e.g., dog or rat) heart, or a specified group of patients. Therefore, the derived elastance curves tend to have limitations for capturing the real-time beat-to-beat and inter-subject variations in $e(t)$ characteristics.

In the present work, certain temporal intervals and amplitudes of the recorded ECG and ECG-derived respiration signals were used to estimate the parameters of $e(t)$ in real time. Such ECG feature combinations are known to be unique to an individual's cardiovascular system [28, 29]. Therefore, the ECG-derived parameters of $e(t)$ tend to capture the effects of physiological differences among individual cardiovascular systems.

4.3 Research approach

Our modeling approach to generating certain surrogate hemodynamic waveforms from measured ECG features is summarized in Figure 4-1. The dataset from PhysioNet's MGH/MF waveform database [30] was used in this paper. We selected 20 recordings (subjects in the 42-84-year age range) in critical care settings with a variety of medical conditions. Each recoding consisted of synchronous measurements of ECG (lead-II), central venous pressure (CVP), pulmonary arterial pressure (PAP), and respiratory impedance (RI) gathered over 12-86 min duration at 360 Hz sampling rate. To reduce the simulation runtime and nonstationarity effects, signals gathered from each subject were partitioned into 60 sec segments. The approach integrates the following four tasks:

(A) *Signal conditioning and feature extraction*: We extracted respiratory components and ECG events (e.g., the peaks of P, R, and T waves and the offset of the T waves) to formulate the activation functions. A phase space reconstruction method was utilized to extract critical events from the measured ECG signal (see Sec. 4.3.1).

(B) *Cardiovascular model formulation*: The model is based on extending Korakianitis and Shi's [31] cardiovascular dynamics model with activation functions obtained from ECG features as excitation inputs (see Sec. 4.3.2).

(C) *Parameter estimation*: The salient model parameters were estimated in real time using a multiple regression model with a compact set of ECG features selected as independent variables. The regression model coefficients were estimated offline (see Sec. 4.3.3).

(D) *Model validation*: We validated the model by comparing the time and frequency patterns of the outputs with those of the recordings from the database (details in Sec. 4.3.4).

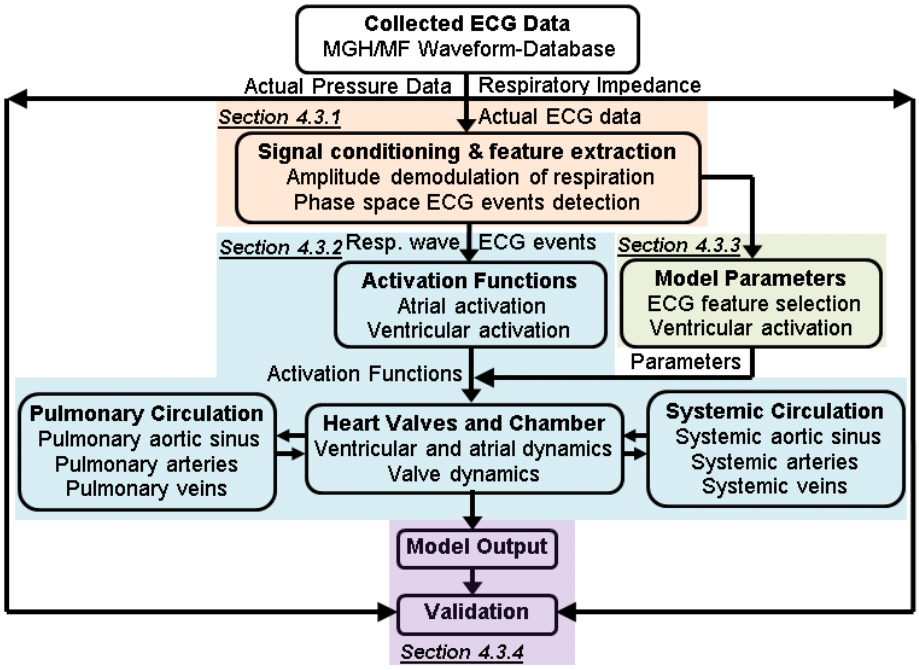


Figure 4-1 Summary of the virtual simulation cardiovascular model

4.3.1 Signal conditioning and feature extraction

4.3.1.1 De-noising and respiration extraction

Low frequency components (0-0.05 Hz) of raw ECG signals are mostly associated with wandering baseline anomalies [32], and high frequencies (>80 Hz) with the noise from the ambient environment, measurement devices, and other artifacts. A band-pass filter with a 0.06-40 Hz passband was employed to remove these extraneous components and retain the ECG features vital for respiration and fiducial point extraction [33]. An amplitude demodulation method [34] was employed to extract the respiration signal $\mathrm{Resp}(t)$ from the ECG. Here, we formed a pulse series with the local average of RR interval values as the pulse intervals; the series was load-modulated by the corresponding R amplitudes. The respiration signal $\mathrm{Resp}(t)$ was extracted from linear interpolation of the resulting pulse series.

4.3.1.2 ECG event detection

First, a wavelet filter [35] was used to detect R peaks in the de-noised ECG signal. Next, a vector time series was obtained by embedding the de-noised ECG signal in a three-dimensional state space with 12 ms time-delay [36]. As shown in Figure 4-2, the trajectories in the reconstructed state space portray three loops−the smallest loop (marked with the solid green line) captures predominantly the P-wave behavior; the largest (the dotted black line), the QRS complex; and the third (the dashed red line) captures the T wave characteristics. The maximum vector magnitudes in the P and T loops were taken to locate the peaks of the P and T waves. The peaks of the Q and S waves were identified at the minimum points in the time domain in PR interval and RT interval, respectively, in the de-noised ECG signal. The J point, defined as the junction of the QRS complex and the ST segment, was designated as the first inflection point (location where ECG waveform changes from concave to convex) after the S peak. The offset of the T wave was set at J + 80 ms if the heart rate (HR) was less than 100 beat/min, J + 72 if 100 ≤

HR < 110, J + 64 if 110 ≤ HR < 120, and J + 60 if HR ≥ 120 [37]. The extracted ECG events thus were used to derive the time profiles of the activation functions and the covariates for a regression model to estimate the model parameters.

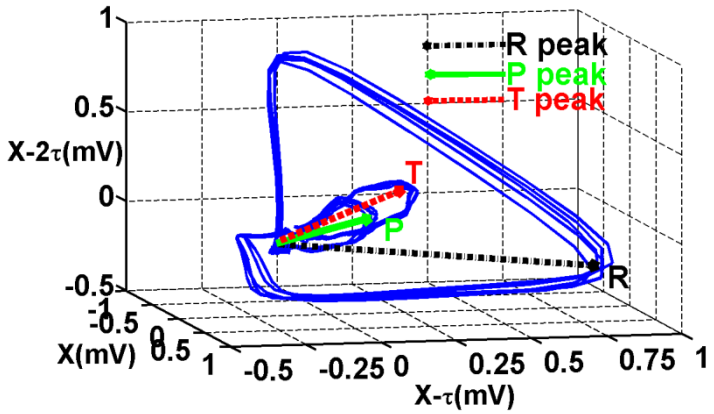


Figure 4-2 Extraction of P, T loop, and QRS complex from VCG

4.3.2 Cardiovascular model formulation

4.3.2.1 Activation functions

The present cardiovascular model uses the elastance characteristics of the four heart chambers and the events extracted from the ECG signal to construct the activation (excitation) functions. Electromechanical delays are neglected because they tend to be a fraction of the signal sampling intervals as well as the time scales of the features (> 0.1 sec) considered in this study. We used the same activation function (u_a) for both atria and an identical activation function (u_v) for both ventricles.

The activation function of the atria (see Figure 4-3), whose support is assumed to be contemporaneous with the atrial systole period, takes the form of the following raised cosine function [38]:

$$u_a(t)=\begin{cases}0.5\times\left(1-\cos\left(\frac{2\pi(t-T_1)}{(T_2-T_1)}\right)\right) & T_1\le t\le T_2\\ 0 & t\notin(T_1,T_2)\end{cases} \quad \text{Eq. 4-1}$$

Here, the onset, T_1, and the offset, T_2, of the atrial systole are marked as the times of P peak and R peak events, respectively.

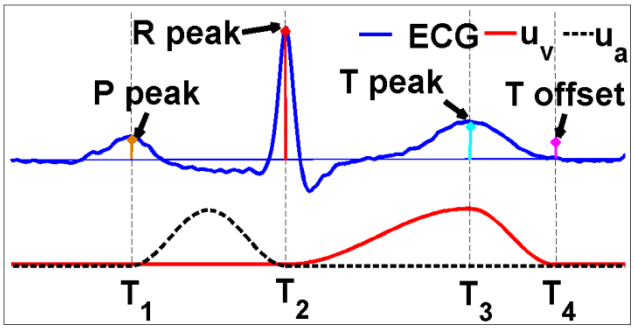


Figure 4-3 Atrioventricular activation functions synthesized from ECG features

Following Chung *et al.*'s findings [39], we used two raised cosine functions to approximate the ventricle activation function during the isovolumic contraction, ejection, and relaxation processes (which take place between the end diastole T_2 and end systole T_4 of a cardiac cycle) as:

$$u_v(t)=\begin{cases}0.5\times\left(1-cos\left(\frac{\pi(t-T_2)}{T_3-T_2}\right)\right) & T_2\le t\le T_3\\ 0.5\times\left(1+cos\left(\frac{\pi(t-T_3)}{T_4-T_3}\right)\right) & T_3<t\le T_4\\ 0 & t\notin(T_2,T_4)\end{cases} \quad \text{Eq. 4-2}$$

where T_2, T_3, and T_4 are assumed to be synchronous with R peak, T peak, and T offset event, respectively (see Figure 4-3). Accordingly, the ventricular activation starts at the R peak, reaches the maximum amplitude at the T peak (the beginning of the ventricular ejection), and subsides at the offset of the T wave. The raised cosine functions used in the present study to approximate activation functions are similar to those used in [31]. Ottesen and Danielsen's [40] comparative study suggested that the activation function shapes based on different premises have very marginal differences.

4.3.2.2 Cardiovascular system

We adapted the model suggested by Korakianitis and Shi [31] to capture the dynamics of the four heart chambers, namely, right and left ventricles and atria and the two systemic and pulmonary circulations of the cardiovascular system. All four sets of equations for the chambers assumed a similar form. For example, the dynamics of left ventricle volume (V_{lv}) and pressure (P_{lv}) are given by:

$$\frac{\mathrm{d}V_{\mathrm{lv}}}{\mathrm{d}t} = \mathrm{CQ}_{\mathrm{ao}}\,\mathrm{N}_{\mathrm{ao}}\sqrt{|P_{\mathrm{lv}}\text{-}P_{\mathrm{sas}}|} \qquad \text{Eq. 4-3}$$

$$P_{\mathrm{lv}} = P_{\mathrm{lv},0} + \mathrm{e}_{\mathrm{lv}}(V_{\mathrm{lv}} - V_{\mathrm{lv},0}) \qquad \text{Eq. 4-4}$$

where N_{ao} depends on the angular position of the aortic valve leaflets θ_{ao}as:

$$N_{\mathrm{ao}} = (1-\cos\theta_{\mathrm{ao}})^2/(1-\cos\theta_{\mathrm{max,ao}})^2 \qquad \text{Eq. 4-5}$$

The elastance functions for the ventricles and atria are expressed in terms of the ECG-derived activation functions as:

$$e_{\mathrm{la}}(t) = E_{\mathrm{lad}} + (E_{\mathrm{las}} - E_{\mathrm{lad}}) \times (\mathrm{u}_{\mathrm{a}}(t) + \gamma\,\mathrm{Resp}(t)) \qquad \text{Eq. 4-6}$$
$$e_{\mathrm{lv}}(t) = E_{\mathrm{lvd}} + (E_{\mathrm{lvs}} - E_{\mathrm{lvd}}) \times (\mathrm{u}_{\mathrm{v}}(t) + \gamma\,\mathrm{Resp}(t))$$

where $E_{\mathrm{lvs}}, E_{\mathrm{lvd}}, E_{\mathrm{las}}, E_{\mathrm{lad}}$ capture the amplitude (gain) of the elastance waveforms, and γ quantifies the coupling strength of the respiratory effect on the cardiovascular system activation beyond constricting and dilating the activation functions in Eqs. (4-1, 4-2) based on changes in heart rate. According to Chung *et al.* [39], ventricular elastance depends on the free wall, pericardium, and thoracic chamber pressures. Assuming the environmental pressure to be invariant and the body movement to be negligible, the intrathoracic chamber pressure increases during expiration, concomitant with the chest/rib cage contraction, and decreases during inspiration, concomitant with the chest expansion. The instantaneous elastance is thus influenced

by respiration. Since the respiration signal $\mathrm{Resp}(t)$ aggregates the aforementioned expansions and contractions, we used the coupling parameter γ to quantify the (first order) sensitivity of the elastance functions to respiration, and hence the thoracic chamber pressure.

Aortic valve dynamics is expressed as [31]:

$$\frac{\mathrm{d}^2\theta_{\mathrm{ao}}}{\mathrm{d}t^2} = (P_{\mathrm{lv}} - P_{\mathrm{sas}})\mathrm{K}_{\mathrm{p,ao}}\cos\theta_{\mathrm{ao}} - \mathrm{K}_{\mathrm{f,ao}}\frac{\mathrm{d}\theta_{\mathrm{ao}}}{\mathrm{d}t} + \mathrm{K}_{\mathrm{b,ao}}Q_{\mathrm{ao}}\cos\theta_{\mathrm{ao}} - \mathrm{K}_{\mathrm{v,ao}}Q_{\mathrm{ao}}(\mu(Q_{\mathrm{ao}})\sin\theta_{\mathrm{ao}} + \mu(-Q_{\mathrm{ao}})\cos\theta_{\mathrm{ao}} \quad \text{Eq. 4-7}$$

where $\mu(.)$ is a unit step function and the remainder of the symbols are listed in the Appendix (Table A1).

The systemic circulation loop consists of the aortic sinus, artery, arteriole, capillary, and vein segments, and each individual component is modeled by considering the local blood flow resistance, the elasticity of blood vessels, and the inertia of blood. The systemic pressures and flow rates are given by:

$$\begin{cases} \frac{\mathrm{dP_{sas}}}{\mathrm{dt}} = \frac{\mathrm{Q_{ao}\text{-}Q_{sas}}}{\mathrm{C_{sas}}}\ ; \frac{\mathrm{dQ_{sas}}}{\mathrm{dt}} = \frac{\mathrm{P_{sas}\text{-}P_{sat}\text{-}R_{sas}Q_{sas}}}{\mathrm{L_{sas}}}; \frac{\mathrm{dP_{sat}}}{\mathrm{dt}} = \frac{\mathrm{Q_{sas}\text{-}Q_{sat}}}{\mathrm{C_{sat}}} \\ \frac{\mathrm{dQ_{sat}}}{\mathrm{dt}} = \frac{\mathrm{P_{sat}\text{-}P_{svn}\text{-}(R_{sat}+R_{sar}+R_{scp})Q_{sas}}}{\mathrm{L_{sat}}}; \frac{\mathrm{dP_{svn}}}{\mathrm{dt}} = \frac{\mathrm{Q_{sat}\text{-}\ Q_{svn}}}{\mathrm{C_{svn}}};\ \mathrm{Q_{svn}} = \frac{\mathrm{P_{svn}\text{-}P_{ra}}}{\mathrm{R_{svn}}} \end{cases} \quad \text{Eq. 4-8}$$

The variation of pressures and flow rates across the pulmonary loop are quantified using similar function forms as detailed by Korakianitis and Shi [31]. Altogether, the cardiovascular process dynamics are expressed in terms of deterministic nonlinear differential equations with the parameters and activation functions estimated from the measured ECG [9].

4.3.3 ECG-based parameter estimation

The model parameters include the elastance characteristics of the left and right ventricles (i.e., E_{lvs}, E_{rvs}, E_{lvd}, and E_{rvd}), the respiratory coupling γ, and six other parameters, namely, $\mathrm{G_{pa}}$, $\mathrm{O_{pa}}$, $\mathrm{G_{ra}}$, $\mathrm{O_{ra}}$, $\mathrm{G_{pv}}$, and $\mathrm{O_{pv}}$, that quantify the gain and offset of the pulmonary arterial pressure

(P_{pa}), right atrial pressure (P_{ra}), and pulmonary venous pressure (P_{pv}) signals. Quantification of these parameters requires complex, invasive measurements taken over several conditions [41].

We developed an offline statistical model that maps these model parameters to appropriate ECG features. The present method is able to adaptively estimate the model parameters and is more responsive to anatomical and physiological characteristics of individual subjects as well as real-time physiological changes as captured in ECG patterns compared to conventional parameter selection methods [4, 42]. Figure 4-4 describes the training (dashed line) and simulation (solid line) phases of the virtual cardiovascular model. The training phase consists of building an empirical multiple regression model to estimate a set of model parameters from the selected ECG features. The simulation phase consists of generating certain hemodynamic signals from the cardiovascular model whose parameters are estimated from online ECG features using the trained regression model.

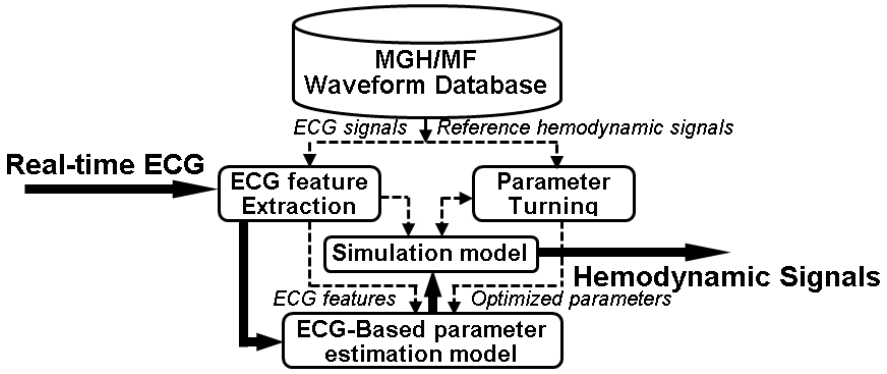


Figure 4-4 Summary of ECG-based parameter estimation method for virtual cardiovascular model

4.3.3.1 Parameter tuning

The purpose of parameter tuning is to determine an optimized model parameter set that minimizes the mean square error between the simulation model outputs and the measurements

from recording signals. The model-estimated pressure signals P_{ra} and P_{pa} are compared with measured central venous pressure (CVP) and pulmonary arterial pressure (PAP) signals, respectively. The vector of tuned parameters consists of G_{pa}, O_{pa}, G_{ra}, O_{ra}, G_{pv}, O_{pv}, γ, E_{lvs}, E_{rvs}, E_{lvd}, and E_{rvd}. A global optimization-based pattern search algorithm *[43]* was used to find the optimized parameters. The optimization problem is defined as follows:

$$\min f(x)_{x\in\Omega} \text{ s.t } l \leq Ax \leq u \qquad \text{Eq. 4-9}$$

where $f: \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ is the objective function, $x\epsilon\boldsymbol{R}^n$, $l, u\epsilon\boldsymbol{R}^m$ and $A\epsilon\mathbf{Q}^{mxn}$. The pattern search algorithm consists of two phases: the search step and the poll step. Initially, x_k is assigned as the starting point and Δ_k is the coarseness of the grid M_k defined over Ω. In the search step, the objective function is evaluated at a finite number of points on the mesh M_k. If $f(x_{k+1}) < f(x_k)$, the search step is considered successful; then the mesh is coarsened to $\Delta_{k+1} \geq \Delta_k$, and the search step is restarted from the improved point x_{k+1}. If an improved point is not found on the mesh, the poll step is invoked by considering the initial points that are the neighbors of x_k on the mesh. Objective functions at the neighboring mesh points are evaluated to see if a lower function value can be found. If an improved mesh point is found (i.e., $f(x_{k+1}) < f(x_k)$), $\Delta_{k+1} > \Delta_k$ is set, and the search is restarted from the improved point. Otherwise, the search is started with mesh refinement $\Delta_{k+1} < \Delta_k$. Multiple (>20) starting points have been used to minimize the probability that the algorithm will converge to a local minimum [44].

4.3.3.2 ECG-based parameter estimation model

We identified the following four groups with a total of 15 ECG features that can collectively track much of the subject-to-subject variations in anatomical and physiological characteristics to estimate the salient model parameters: (i) sample averages of P, QRS, T peak amplitudes, (ii) sample averages of RR, PR, ST, QT intervals, (iii) standard deviations and differences between intervals, and (iv) the area swept by the ST segment. These features were selected based on Israel

et al.'s [29] finding that an ECG trace expresses cardiac features unique to an individual. From a physiological standpoint, changes in R and T amplitudes in frontal and precordial leads are associated with instantaneous changes in the volume and output of the left ventricle [45, 46]. Specifically, the R amplitude is sensitive to the radial movement of the heart (dilation) in relation to the chest wall and the T wave amplitude to the variations in the ratio of the endocardial to the epicardial surface area of the left ventricle.

The heart rate (estimated from RR interval) is determined by the vagal-sympathetic mechanisms, and the systolic period (estimated from intra-beat intervals of ECG) is affected by the sympathetic efferent discharge frequency (F_{con}), which in turn influences the peak elastance [42]. A rise in F_{con} increases the maximum elastance and shortens the ventricular systole. Also, according to Akselrod *et al.* [47], short term fluctuations of heart rates are affected by the autonomic control levels at pacemaker sites. Pagani *et al.* [48] showed that low frequency heart rate fluctuations provide an index of sympathetic efferent activity, which are associated with the peak ventricular elastance. The standard deviations of the ECG intervals (RR, ST, PR, QT intervals) can therefore be used to capture the effects of vagal and sympathetic neural activities, and hence contribute to the estimation of the peak ventricular elastance.

The selected features, extracted via phase space analysis described in Section 4.3.1.2, were used as predictors for a regression model to derive model parameters. It may be noted that many ECG features contain redundant information, and each model parameter tends to be sensitive to a different combination of features. We used principal component analysis (PCA) to address these redundant and diverse relationships. The subsets of ECG features whose contributions to the principal components are the largest were selected for parameter estimation.

To predict the parameters of the cardiovascular model, we considered a multiple linear regression model $\Phi = X\beta + \epsilon$, where Φ is an n x 1 vector of model parameters, X is an n x p full-column rank matrix of predictors (i.e., ECG features), β is a p x 1 vector of unknown regression model coefficients, and ϵ is an independent Gaussian random variable. A backward stepwise procedure [49] was used to select the most predictive feature combinations. The procedure begins with considering all variables in the model and sequentially deleting one variable at a time. A partial F-test for each variable in the presence of the others is conducted with the test statistic $F = \frac{(\hat{\beta}X^T\Phi - n\bar{\Phi}^2)/p}{\Phi^T\Phi - \hat{\beta}X^T\Phi/(n-p-1)} \sim F_{p,n-p-1}$. The stepwise procedure continues until the smallest F has p-value > 0.05. A k-fold cross validation was used to assess for the generalizability of the regression model. A random 90% of the dataset was selected for fitting the multiple regression model and the remaining 10% for validation. This process was repeated 20 times to assess the consistency of the regression coefficients.

4.3.4 Model validation

While the model-generated pressure waveforms can be compared with those from the actual recordings, only certain extreme values of the waveforms are considered clinically important [50, 51]. For example, many of the clinical assessment procedures for a subsequent coronary heart disorder use the extrema (systolic and diastolic) blood pressure waveforms [52]. Therefore, one needs a test procedure that weights the waveform conformance according to the clinical importance of the specified portion or pattern of the waveform. In this study, the Anderson-Darling goodness of fit test [53] was used to measure the similarity of the systolic and diastolic pressures between the model outputs and the actual measurements. A two-sample Anderson-Darling statistic is defined as:

$$A_{nm}^2 = \frac{nm}{N}\int_{-\infty}^{\infty}\frac{\{F_n(x)-G_m(x)\}^2}{H_N(x)\{1-H_N(x)\}}dH_N(x) \quad \text{Eq. 4-10}$$

$$H_N(x) = [nF_n(x) + mG_m(x)]/N \quad \text{Eq. 4-11}$$

where $F_n(x)$ and $G_m(x)$ are the empirical distribution functions of the measured (n samples) and model-derived (m samples) pressure (systolic and diastolic) signals, respectively, and $H_N(x)$ is the empirical distribution function of the combined sample size an $N = n + m$. The distinct values in the combined dataset, ordered from smallest to largest, are denoted as $z_1, z_2, \ldots z_N$. Here, we aimed to compare the systolic and diastolic blood pressure values extracted from the model waveform to those from the measurements. As stated in the foregoing, the Anderson-Darling statistic is more appropriate for the present context than the other two-sample statistics (e.g., K-S test, Cramer-von Mises) because it places more weight on observations in the tails of a distribution [53, 54] through the use of the weight function $[H(x)(1-H(x))]^{-1}$ in Eq. 4-10. The test statistic can be estimated from a sample (z_j) as:

$$A^2 = \frac{n+m-1}{(n+m)^2}\times\left[\frac{1}{n}\sum_{j=1}^{N}\frac{h_j((n+m)F_j-nH_j)^2}{H_j(n+m-H_j)-\frac{(n+m)h_j}{4}}+\frac{1}{m}\sum_{j=1}^{N}\frac{h_j((n+m)G_j-mH_j)^2}{H_j(n+m-H_j)-\frac{(n+m)h_j}{4}}\right] \quad \text{Eq. 4-12}$$

where h_j is the number of observations of the combined sample equal to z_j, H_j is the sum of the number of samples combined samples with values less than z_j plus one half the number of values in the combined samples equal to z_j, and F_j and G_j are the sums of the number of samples from the measured and model-derived signals, respectively, that are less than z_j plus one half the number of values in that group that are equal to z_j.

Under the null hypothesis that $F(x)$ and $G(x)$ are drawn from the same distribution, the variance of the test statistics A^2 in Eq. 4-12 is given by:

$$\sigma^2 = var(A^2) = \frac{aN^3+bN^2+cN+d}{(N-1)(N-2)(N-3)}, \quad \text{Eq. 4-13}$$

where a, b, c, d are derived from [54]. If the test statistic A^2 in (13) is above a critical value given by $ADC = 1 + 1.961\sigma$, then one can reject (at significance level α=0.05) the hypothesis that the measurements and the model-derived signals were drawn from the same distribution.

In addition, Kullback-Leibler (KL) divergence [55] was used to assess the closeness of the systolic and diastolic pressure distributions of the model outputs to the actual measurements. KL divergence is given by:

$$D_{KL}(g; f) = \sum_{i=1}^{k} g(x_i) log\{g(x_i)/f(x_i)\}, \quad \text{Eq. 4-14}$$

where f, g are the density functions of the measured and the model-derived signal, respectively. As KL divergence $D_{KL}(g; f)$ is an asymmetric measurement, we used $M_{KL} = [D_{KL}(g; f) + D_{KL}(f; g)]/2$ as a symmetrized metric to compare the distribution similarity. The larger the KL divergence value, the farther apart are the two signals [55]. The KL divergence value equals zero when the two distributions are identical.

4.4 Implementation details and results

The model was implemented in the Matlab/Simulink environment. The sample times of the signal sources and the model were matched so that the outputs of the model were synchronized with the input ECG signal. Every simulation was run for 60 sec. The transient effects were noted to subside by the end of 10 sec. Therefore, data collected after 10 sec of the simulation was used for further analysis. The model outputs consist of waveforms of the pressure, volume, flow rate from the four heart chambers, and the pulmonary and systemic circulation modules.

PCA was employed to reduce the statistical redundancy between the high dimensional ECG features without significant loss of information. The application of PCA suggested that the first

four principal components could explain approximately 78% of the variations in the data, and thus these four were considered for further studies. None of the other components captured more than 5% of the variation in the feature values. Table 4-1 shows the descriptions of 15 features and the contributions of each feature in four principal eigen directions. The larger the coefficients, the higher the contribution of that feature to the variation along that eigen direction. It is evident that all except the RR standard deviation (RR Std) and the difference ST−PR contributed to one of the leading components. The remaining 13 significant ECG features were used as predictors in the training and simulation phases.

Table 4-1 Contribution of ECG features to the first four principal components

Feature	Description (Unit)	Comp 1	Comp 2	Comp 3	Comp 4
R AmpAvg	Average of R peak amplitudes (mV)	-0.03	0.36	-0.10	0.27
P AmpAvg	Average of P peak amplitudes (mV)	-0.26	0.26	0.00	-0.41
T AmpAvg	Average of T peak amplitudes (mV	-0.10	0.35	-0.08	0.37
RR Avg	Average of RR intervals (s)	0.31	0.15	-0.25	-0.01
PR Avg	Average of PR intervals (s)	-0.10	0.12	-0.04	-0.45
ST Avg	Average of ST intervals (s)	0.42	0.05	0.07	-0.21
QT Avg	Average of QR intervals (s)	0.42	0.03	0.11	-0.24
RR Std	Standard deviation of RR intervals (s)	-0.03	0.16	0.09	-0.27
PR Std	Standard deviation of PR intervals (s)	0.28	0.17	-0.12	0.40
ST Std	Standard deviation of ST intervals (s)	0.05	-0.11	-0.65	-0.11
QT Std	Standard deviation of QT intervals (s)	0.05	-0.10	-0.65	-0.11
ST-PR	ST, PR interval differences (s)	0.23	-0.02	0.08	0.05
QT-PR	QT, PR interval differences (s)	0.44	-0.02	0.12	-0.01
ST Area	Average of ST segment area (mV s)	-0.01	0.54	0.05	-0.02
STAreaSd	Stand. Dev. of ST segment area (mV s)	0.06	0.52	-0.10	-0.05

In the training phase, the coefficient vectors of the empirical regression model were estimated offline. First, the pattern search (optimization) method described in Section 4.3.3.1 was used to tune the parameters pertaining to the elastance (E_{lvs}, E_{lvd}, E_{rvs}, and E_{rvd}), the offsets and gains of pressures (G_{pa}, O_{pa}, G_{ra}, O_{ra}, G_{pv}, and O_{pv}), and the respiration coupling γ. The tuned parameters were regressed offline with selected ECG features to estimate the empirical regression model. Table 4-2 lists the significant ECG features (with p-value < 10^{-3}) and the corresponding coefficients of the empirical regression model. The subsequent simulation phase involved

generating the hemodynamic signals in real time from the parameterized model. The model parameters in the simulation phase were estimated using the regression model with the selected ECG features as predictors. The values of the various parameters used in the model are listed in the Appendix (Table A2).

Next, we compared the salient hemodynamic signals extracted from the model outputs with the actual measurements in the database. The following combinations of signals were compared:

i) Chamber pressures and volumes with ideal profiles [56].

ii) Right atrial pressure (P_{ra}) from the model with the measured central venous pressure (CVP).

iii) Pulmonary arterial pressures from the model (P_{pa}) with those from actual measurements (PAP).

iv) Pulmonary vein pressure from the model (P_{pv}) with the measured respiratory impedance (RI).

v) Systolic and diastolic (max and min) pressure values estimated from the model versus measurements.

Table 4-2 Coefficients of regression model to estimate model parameters from ECG features

Feat. / Par	Intercept	R AmpAvg	P AmpAvg	T AmpAvg	RR Avg	PR Avg	ST Avg	QT Avg	ST Std	PR Std	QT Std	QT-PR	ST Area	ST AreaSd
G_{pa}	**3.24**						**0.3**		**-0.22**	**0.7**				
O_{pa}	**2.35**	**0.52**		**-0.1**	**0.68**									
G_{ra}	**0.4**			**0.1**				**-0.21**					**0.3**	**-0.14**
O_{ra}	**-0.3**					**0.25**	**-0.18**							
G_{pv}	**4.35**			**0.91**				**0.12**	**0.25**				**-3.23**	
O_{pv}	**10.51**	**1.14**				**-0.26**						**-3.16**		
γ	**1.32**	**-2.8**	**0.12**							**0.45**			**1.37**	**-1.48**
E_{lvs}	**0.94**		**0.25**	**0.12**			**-1.15**	**1.35**					**0.25**	**0.25**
E_{rvs}	**0.92**		**0.2**			**3.19**		**-3.29**			**-0.97**		**1.72**	
E_{lvd}	**0.12**		**0.19**						**2.41**	**-2.31**		**-0.35**		**0.41**
E_{rvd}	**0.09**	**0.15**							**0.14**			**-0.13**		

4.4.1 Pressure and volume waveforms

The various patterns of the model-generated waveforms of pressures (Figure 4-5) and volumes (Figure 4-6) of the four heart chambers lay within similar ranges of values as those of ideal profiles [56]. For the left ventricle, the pressure ranged from 0 to 125 mmHg, the volume varied between 70-130 ml, and the shapes (e.g., skewness and support) of the pressure and the volume waveform patterns are also comparable with those of the ideal profiles [56].

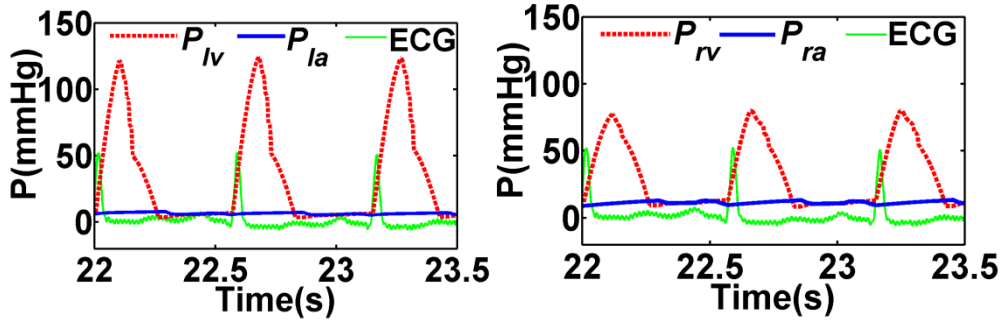


Figure 4-5 Waveform of pressure (P) of left atrium and left ventricle (left), right atrium and right ventricle (right)

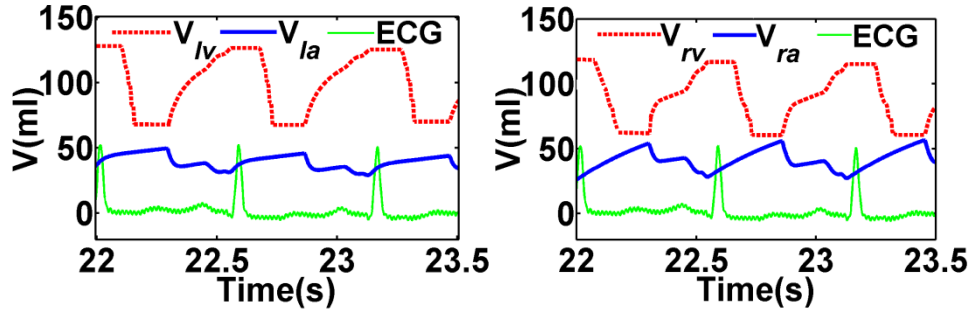


Figure 4-6 Waveform of volume (V) of left atrium and left ventricle (left), right atrium and right ventricle (right)

4.4.2 Pulmonary arterial pressure comparisons

Figure 4-7 shows the measured pulmonary arterial pressure (PAP) and the model output (P_{pa}) waveforms in time (left) and frequency (right) domains. In the time domain, the pulmonary arterial pressure values range from 10mHg through 30mHg for both measured and model-generated waveforms and exhibit similar patterns, including the skewness and rise and drop rates.

The frequency spectrum portrait shows the presence of respiratory and heart rate components at 0.15 Hz, and 0.93 Hz, respectively for both measured and model-derived data. Since PAP is measured at the phlebostatic axis, which is found at the intersection of the midaxillary line and a line drawn from the fourth intercostal space at the right side of the sternum on the thorax, the respiration is also included. These similarities between the pulmonary arterial pressure waveforms suggest that the model-generated P_{pa} signals can be a suitable surrogate for PAP.

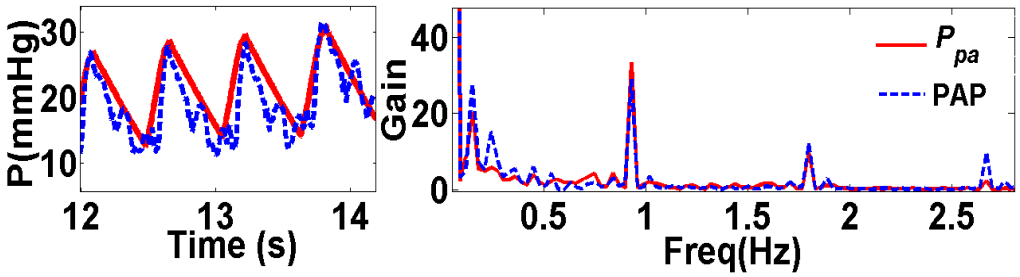


Figure 4-7 Comparison of (left) time and (right) frequency portraits of model derived (solid red line) and measured (dashed blue line) pulmonary arterial pressures

4.4.3 Right atrial pressure and central venous pressure comparisons

The model-generated right atrial pressure (P_{ra}) waveform was compared with the measured central venous pressure (CVP) waveform. Since CVP captures the blood pressure in the thoracic vena cava near the atrium of the heart, it is considered a reasonable surrogate of P_{ra} [56]. Figure 4-8 shows the variation of CVP and P_{ra} in time (left) and frequency (right) domains. It may be noted that the P_{ra} waveform captures the amplitudes, and the time-domain patterns including the skewness and rise and drop rates of the CVP waveform. The frequency portraits were almost identical for both waveforms, with the frequency peaks of respiratory components at 0.15 and 0.24 Hz, and heart-rate components at 0.93 Hz. The similarities between the signal waveforms in both the time and frequency suggest that P_{ra} derived from the model output may serve as a viable surrogate for CVP.

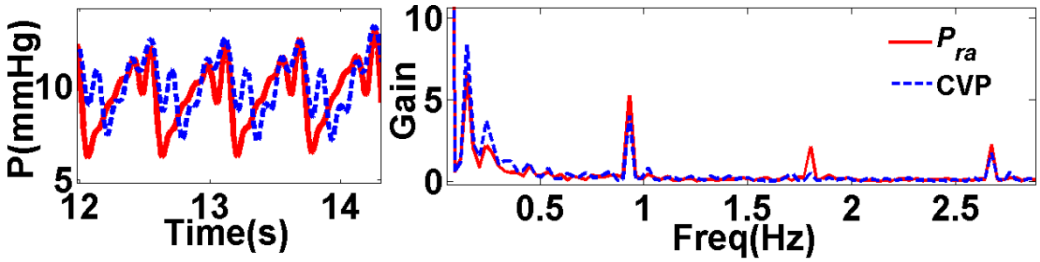


Figure 4-8 Comparison of (left) time and (right) frequency portraits of model-derived right atrial pressure (P_{ra}, solid red line) and the measured central venous pressure (CVP, dashed blue line)

4.4.4 Pulmonary vein pressure and respiratory impedance comparisons

A comparison of the waveforms of the low frequency (≤ 0.3 Hz) component of the pulmonary vein pressure (P_{pv}) and the respiratory impedance (RI) measurement in the time domain is shown in Figure 4-9. The time-domain patterns contained in the low frequency component of the P_{pv} waveform, including the respiratory components, are consistent with those of the RI, the change in the chest movement during the respiratory process. The P_{pv} captures the pressure of blood returning from the lung to the left atrium of the heart. As the elastance of the thorax can be assumed to be constant, the low frequency of the pulmonary venous pressure can be used for the measurement of the respiratory volume [57, 58]. Evident from the figure is that the low frequency components of the model-generated P_{pv} signal can be used to approximate the RI measurement. However, further analyses and calibration of the elastance of the thorax are needed to estimate the actual values of the respiratory signals.

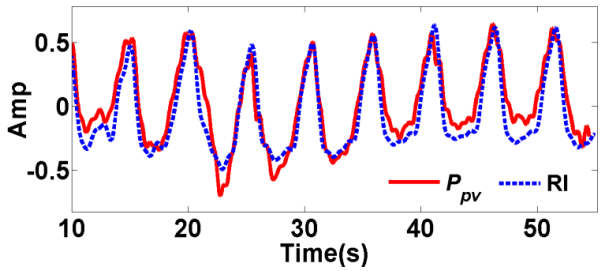


Figure 4-9 Comparison of the waveforms of the model-derived pulmonary venous pressure P_{pv} and the measured respiratory impedance RI

Correlation coefficients ρ and R^2 statistic values were calculated to quantitatively assess the correlations between the model-derived and measured signals (P_{pa} vs. PAP, P_{ra} vs. CVP, and P_{pv} vs. RI). A 10-fold cross validation was used to evaluate the accuracy of the model with an independent dataset. First, we randomly partitioned the dataset into 10 subsets, one of which was used for testing and other nine to build the parameter regression model. For model validation, the ECG features of the testing subset were used to estimate the model parameters. This process of partitioning followed by training and testing was repeated 20 times. Table 4-3 summarizes the average ρ and R^2 values of the various model-derived vs. measured waveforms. It is observed that the P_{pv} waveform matches well with the RI's with $\rho = 0.77$ and $R^2 = 0.79$. The correlations between other model-derived signals and the actual measurements are substantial ($\rho \geq 0.68$, and $R^2 \geq 0.70$).

Table 4-3 Model-derived vs. measured waveform comparisons

Signal waveform comparisons	ρ	R^2
P_{pa} vs. PAP	0.71	0.73
P_{ra} vs. CVP	0.68	0.70
P_{pv} vs. RI	0.77	0.79

4.4.5 Systolic and diastolic pressure comparisons

A two-sample Anderson-Darling goodness-of-fit hypothesis test and KL divergence were used to evaluate the similarity of the systolic and diastolic values from the model outputs with those from the measured pressure profiles. During each heartbeat, the blood pressure varies between the maximum (systolic) and minimum (diastolic) values. Four sets of systolic (upper tail) and diastolic (lower tail) pressure samples were obtained from the pressure value distribution, each obtained by considering the 4th, 5th, 6th, or the 7th percentile (4-7% of) the observed signal realizations. The choice of extreme values (4-7%) was determined based on prior results [59] that suggested that the augmentation (extreme) portions of an atrial pressure waveform are roughly

5% of the pulse pressure (i.e., waveform amplitude) for heart rates in the 60-110 beat/min range. Table 4-4 summarizes the rejection rate (at significance level $\alpha = 0.05$) of the Anderson-Darling test with the null hypothesis that the tails of the model-derived pressure waveforms and those of the measured pressure waveforms emerge from the same distributions. It is noted that for most of the comparisons the rejection rates are below 10%. These low rejection rates, in the 4-7% cut-off range, further indicate that the systolic and diastolic pressures of PAP and CVP can be statistically captured by the model-generated P_{pa} and P_{ra} waveforms, respectively.

Table 4-4 Average rejection rates from Anderson-Darling test

Extreme value cut-off	Pulmonary Arterial Pressure vs. Pulmonary Arterial Pressure		Central Venous Pressure vs. Right Atrial Pressure	
	Systole	Diastole	Systole	Diastole
4%	5%	5%	5%	10%
5%	0	0%	5%	10%
6%	5%	5%	5%	10%
7%	10%	10%	10%	15%

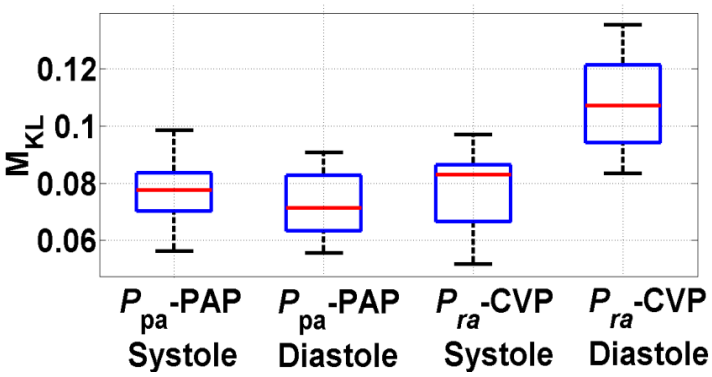


Figure 4-10 KL convergence variations of systole and diastole pressures between model outputs and actual measurements

Next, KL information (M_{KL}) was used to quantify the difference between the distributions of the extrema of the pressure waveforms from the model outputs and those from the measurements. Figure 4-10 shows the variation of the M_{KL} of the systolic and diastolic pressures of the model-

derived signals and those of the actual measurements at a threshold of 5%. It is noted that the average KL information of the model-generated systolic and diastolic samples vs. those from recorded pressures are below 0.14. Such low values of KL information suggest that systolic and diastolic pressures from the model are comparable to those from actual measurements [55].

4.5 Conclusions

This paper reports an approach to generating multiple synchronized hemodynamic signals from cardiovascular systems in real time using ECG features. The research is among initial efforts to use ECG to formulate activation functions for cardiovascular dynamic models. ECG features are used to construct atrioventricular excitation inputs to a nonlinear deterministic lumped parameter model of cardiovascular system dynamics. Traditional approaches to acquiring these hemodynamic signals, such as atrial blood pressure, cardiac output, or the opening angles of the heart valves, require invasive and/or expensive medical instruments [60, 61]. Important events of the ECG signal are extracted by using wavelet analysis and a phase space method. Respiration components extracted from ECG play an important role in the representation of the model outputs. It is noteworthy that the model outputs were correlated substantially with the measured signals when the activation function included the respiration components. The salient time and frequency patterns of the model outputs, including right atrial pressure (P_{ra}), pulmonary arterial pressure (P_{pa}), and the low frequency components of pulmonary vein pressure (P_{pv}) were statistically consistent with those of actual recordings. The waveforms of P_{pa}, P_{ra} and P_{pv} were substantially correlated with those of PAP, CVP and RI with R^2 of 0.73, 0.7, and 0.79, respectively. In addition, the results from a two-sample Anderson-Darling hypothesis test and KL information comparisons suggest strong statistical similarities between the distributions of model-derived and measured diastolic and systolic pressure values.

The similarities of the time and frequency domain characteristics as well as the statistically indistinguishable distribution of the tails of the model outputs and the measured waveforms indicate the suitability of using ECG features to form the excitation inputs for the cardiovascular model. Apart from the accuracy of the model outputs, the causal relationship captured in the model between the patterns of ECG and certain hemodynamic signals can provide valuable information for diagnosis because certain pathologies not easy to diagnose from the ECG patterns may be enhanced in the model outputs, such as pressure, volume, and blood flow rate signals. Further, the model provides an efficient tool for quantitatively assessing the underlying couplings between the mechanical and electrical components of heart dynamics and an ECG. It can also be used to simulate certain hemodynamic signals under specific cardiac disorders for each individual heart and to analyze the dependencies among ECG and other signal features and pathologies. Taken together, the results point to the viability of a virtual cardiovascular instrument platform where the derived hemodynamic signals can be used for clinical diagnosis and treatment. Such a platform can offer the advantages of cost efficiency and time savings.

Pertinently, the current model is limited by the use of the ECG alone to extract multiple hemodynamic signals. The ECG recordings may not effectively capture certain aspects of cardiovascular dynamics, especially those related to systemic and pulmonary circulations that determine the patterns of various hemodynamic signals. Furthermore, the model uses highly simplified expressions to model the effects of vascular components of cardiovascular regulation and hemodynamics and mostly ignores the effects of the nervous system and the kidney in rapid and long term control of arterial pressures. The model also does not contain explicit descriptions of the neural control scheme (e.g., autonomic nervous control of heat rate, myocardial contractility and vasomotor tone), hormones, volume receptors, and metabolism that might play important roles in characterizing the cardiovascular system dynamics. Since the dynamics of the

cardiovascular system are characterized as a lumped parameter model where the model components (e.g., heart chambers, heart valves, arteries, arterioles, capillaries, and vein segments) are treated as homogeneous elements, the model is not applicable in modeling the heart with localized disorders. To further improve the model's accuracy, it is necessary to improve the structures of the dynamics model by involving other important effects of the circulatory hemodynamics, pulmonary mechanics, and ventilator control, as well as using more generalized regression models for parameter estimation. Also, the suitability of the model to characterize the complexity of the human cardiovascular system (e.g., homeostatic processes, adaptive control of heart rate, and varying degrees of contractility) and to capture the dynamics of larger sets of individuals (e.g., age groups, genders, disorders, and medication conditions) remains to be investigated. A more comprehensive visualization and quantification of the heart, hemodynamic couplings, and their relationship with ECG and other signal features can lead to more responsive and cost-effective healthcare delivery.

APPENDIX

Table A1 Nomenclature and subscripts

Nomenclature		**Subscripts (cont')**	
C	compliance	ao	aortic valve
e	elastance	mi	mitral valve
K,k	coefficient	po	pulmonary aortic valve
M	mass	ti	tricuspid valve
P	pressure	pas	pulmonary aortic sinus
Q	flow rate	par	pulmonary arterioles
V	volume	pa	pulmonary arterial blood
θ	phase angle	pcv	pulmonary capillary
Subscripts		pv	pulmonary vein
0	initial value, offset value.	sas	systemic aortic sinus
1,2,...	subscript state variables	sar	systemic arterioles
s	systole	sat	systemic arterial blood
d	diastole	scv	systemic capillary
lv	left ventricle	svn	systemic vein
rv	right ventricle		
la	left atrium		
ra	right atrium		

Table A2 Cardiovascular system parameters

Left heart			**Right heart**		
Par.	***Init. Val.***	***Unit***	***Par.***	***Init. Val.***	***Unit***
α	0.2		α	0.2	
CQ_{ao}	350	ml/(s $mmHg^{0.5}$)	CQ_{po}	350	ml/(s $mmHg^{0.5}$)
CQ_m	400	ml/(s $mmHg^{0.5}$)	CQ_{ti}	400	ml/(s $mmHg^{0.5}$)
E_{lvs}	2.5	mmHg/ml	E_{rvs}	1.15	mmHg/ml
E_{lvd}	0.1	mmHg/ml	E_{rvd}	0.1	mmHg/ml
$P_{lv,0}$	1	mmHg	$P_{rv,0}$	1	mmHg
$V_{lv,0}$	5	ml	$V_{rv,0}$	10	ml
E_{las}	0.25	mmHg/ml	E_{ras}	0.25	mmHg/ml
E_{lad}	0.15	mmHg/ml	E_{rad}	0.15	mmHg/ml
$P_{la,0}$	1	mmHg	$P_{ra,0}$	1	mmHg
$V_{la,0}$	4	ml	$V_{ra,0}$	4	ml
Systemic Circulation			**Pulmonary Circulation**		
Par.	***Init. Val.***	***Unit***	***Par.***	***Init. Val.***	***Unit***
C_{sas}	0.08	ml/mmHg	C_{pas}	0.18	ml/mmHg
R_{sas}	0.003	mmHg s/ml	R_{pas}	0.002	mmHg s/ml
L_{sas}	$62x10^{-6}$	mmHg s^2/ml	L_{pas}	$52x10^{-6}$	mmHg s^2/ml
C_{sat}	1.6	ml/mmHg	C_{pa}	3.8	ml/mmHg
R_{sat}	0.05	mmHg s/ml	R_{pa}	0.01	mmHg s/ml
L_{sat}	0.0017	mmHg s^2/ml	L_{pa}	0.0017	mmHg s^2/ml
R_{sar}	0.5	mmHg s/ml	R_{par}	0.05	mmHg s/ml
R_{scp}	0.52	mmHg s/ml	R_{pcp}	0.25	mmHg s/ml
R_{svn}	0.075	mmHg s/ml	R_{pv}	0.006	mmHg s/ml
C_{svn}	20.5	ml/mmHg	C_{pv}	20.5	ml/mmHg
C_{svc}	1.5	ml/mmHg	C_{pvc}	1.5	ml/mmHg
V_{lv0}	800	ml	V_{rv0}	500	ml
Aortic and Mitral Valve			**Pulmonary and tricuspid Valves**		
Par.	***Init. Val.***	***Unit***	***Par.***	***Init.Val.***	***Unit***
$K_{p,ao}$	5500	rad/(s m)	$K_{p,po}$	5500	rad/(s m)
$K_{f,ao}$	50	1/s	$K_{f,po}$	50	1/s
$K_{b,ao}$	2	rad/(s m)	$K_{b,po}$	2	rad/(s m)
$K_{v,ao}$	7	rad/(s m)	$K_{v,po}$	3.5	rad/(s m)
$K_{p,mi}$	5500	rad/(s m)	$K_{p,ti}$	5500	rad/(s m)
$K_{f,mi}$	50	1/s	$K_{f,ti}$	50	1/s
$K_{b,mi}$	2	rad/(s m)	$K_{b,ti}$	2	rad/(s m)
$K_{v,mi}$	3.5	rad/(s m)	$K_{v,ti}$	3.5	rad/(s m)

REFERENCES

[1] V. L. Roger, *et al.*, "Heart Disease and Stroke Statistics—2011 Update " *Circulation,* vol. 123, e18-e209, 2011.

[2] NHLBI. (2009). *Morbility and Mortality: 2009 Chart Book on Cardiovascular, Lung, and Blood Diseases*. Available: http://www.nhlbi.nih.gov/resources/docs/2009_ChartBook.pdf

[3] G. Pennati, *et al.*, "Mathematical Modelling of the Human Foetal Cardiovascular System Based on Doppler Ultrasound Data," *Medical Engineering & Physics,* vol. 19, 327-335, 1997.

[4] T. Heldt, *et al.*, "Computational Modeling of Cardiovascular Response to Orthostatic Stress," *Journal of Applied Physiology,* vol. 92, 1239-1254, March 1, 2002.

[5] A. Quarteroni and L. Formaggia, "Mathematical Modelling and Numerical Simulation of the Cardiovascular System," in *Handbook of Numerical Analysis*. vol. Volume 12, N. Ayache, Ed., ed: Elsevier, 2004, pp. 3-127.

[6] J. Wan, *et al.*, "A One-Dimensional Finite Element Method for Simulation-Based Medical Planning for Cardiovascular Disease," *Computer Methods in Biomechanics and Biomedical Engineering,* vol. 5, 195-206, 2002.

[7] F. N. Van De Vosse, *et al.*, "Finite-Element-Based Computational Methods for Cardiovascular Fluid-Structure Interaction," *Journal of Engineering Mathematics,* vol. 47, 335-368, 2003.

[8] R. Kerckhoffs, *et al.*, "Coupling of a 3d Finite Element Model of Cardiac Ventricular Mechanics to Lumped Systems Models of the Systemic and Pulmonic Circulation," *Annals of Biomedical Engineering,* vol. 35, 1-18, 2007.

[9] T. Q. Le, *et al.*, "Towards Virtual Instruments for Cardiovascular Healthcare: Real-Time Modeling of Cardiovascular Dynamics Using Ecg Signals," in *IEEE Conference on Automation Science and Engineering (CASE)*, 2010, pp. 903-910.

[10] A. Daneels and W. Salter, "What Is Scada?," in *International Conference on Accelerator and Large Experimental Physics Control Systems*, 1999.

[11] R. Pearse, *et al.*, "Equipment Review: An Appraisal of the Lidcotmplus Method of Measuring Cardiac Output," *Critical Care,* vol. 8, 190 - 195, 2004.

[12] J. Smith, *et al.*, "Monitoring Arterial Blood Pressure and Cardiac Output Using Central or Peripheral Arterial Pressure Waveforms," J.-L. Vincent, Ed., ed Berlin: Springer 2009, pp. 285-296.

[13] A. Noordergraaf, *et al.*, "The Use of an Analog Computer in a Circulation Model," *Progress in Cardiovascular Diseases - Elsevier,* 419-439, 1963.

[14] G. N. Jager, *et al.*, "Oscillatory Flow Impedance in Electrical Analog of Arterial System: Representation of Sleeve Effect and Non-Newtonian Properties of Blood," *Circulation Research,* vol. 16, 121-133, 1965.

[15] A. Avolio, "Multi-Branched Model of the Human Arterial System," *Medical and Biological Engineering and Computing,* vol. 18, 709-718, 1980.

[16] V. Rideout and D. Dick, "Difference-Differential Equations for Fluid Flow in Distensible Tubes," *Physics in Medicine and Biology,* 178–187, 1967.

[17] A. C. Guyton, *et al.*, "Circulation: Overall Regulation," *Annual Review of Physiology,* vol. 34, 13-44, 1972.

[18] K. E. Simanonok, *et al.*, "A Comprehensive Guyton Model Analysis of Physiologic Responses to Preadapting the Blood Volume as a Countermeasure to Fluid Shifts," *Journal of Clinical Pharmacology,* vol. 34, 440-453, 1994.

[19] J. Werner, *et al.*, "Simulation and Prediction of Cardiotherapeutical Phenomena from a Pulsatile Model Coupled to the Guyton Circulatory Model," *IEEE Transactions on Biomedical Engineering,* vol. 49, 430-439, 2002.

[20] R. Zhang, *et al.*, "Arterial-Cardiac Baroreflex Function: Insights from Repeated Squat-Stand Maneuvers," *American Journal of Physiology Heart and Circulation Physiology,* vol. 297, R116-123, July 1, 2009.

[21] R. Beyar, *et al.*, "Interaction between Cardiac Chambers and Thoracic Pressure in Intact Circulation," *American Journal of Physiology - Heart and Circulatory Physiology,* vol. 253, H1240-H1252, 1987.

[22] F. Liang and H. Liu, "A Closed-Loop Lumped Parameter Computational Model for Human Cardiovascular System," *International Journal Series C Mechanical Systems, Machine Elements and Manufacturing,* vol. 48, 484-493, 2005.

[23] H. Suga, *et al.*, "Load Independence of the Instantaneous Pressure-Volume Ratio of the Canine Left Ventricle and Effects of Epinephrine and Heart Rate on the Ratio," *Circulation Research,* vol. 32, 314-322, 1973.

[24] D. Burkhoff, *et al.*, "Contractility-Dependent Curvilinearity of End-Systolic Pressure-Volume Relations," *American Journal of Physiology Heart and Circulation Physiology,* vol. 252, H1218-1227, June 1, 1987.

[25] D. Kass, *et al.*, "Influence of Contractile State on Curvilinearity of in Situ End-Systolic Pressure-Volume Relations," *Circulation,* vol. 79, 167-178, 1989.

[26] H. Senzaki, *et al.*, "Single-Beat Estimation of End-Systolic Pressure-Volume Relation in Humans: A New Method with the Potential for Noninvasive Application," *Circulation,* vol. 94, 2497-2506, 1996.

[27] S. Klotz, *et al.*, "Single-Beat Estimation of End-Diastolic Pressure-Volume Relationship: A Novel Method with Potential for Noninvasive Application," *American Journal of Physiology Heart and Circulation Physiology,* vol. 291, H403-412, July 1, 2006.

[28] J. M. Irvine, *et al.*, "Eigenpulse: Robust Human Identification from Cardiovascular Function," *Pattern Recognition,* vol. 41, 3427-3435, 2008.

[29] S. A. Israel, *et al.*, "Ecg to Identify Individuals," *Pattern Recognition,* vol. 38, 133-142, 2005.

[30] A. L. Goldberger, *et al.*, "Physiobank, Physiotoolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation,* vol. 101, E215-20, Jun 13 2000.

[31] T. Korakianitis and Y. Shi, "Numerical Simulation of Cardiovascular Dynamics with Healthy and Diseased Heart Valves," *Journal of Biomechanics,* vol. 39, 1964-1982, 2006.

[32] E. Fitzgibbon, *et al.*, "Determination of the Noise Source in the Electrocardiogram During Cardiopulmonary Resuscitation," *Critical Care Medicine,* vol. 30, S148-S153, 2002.

[33] S. Bukkapatnam, *et al.*, "Classification of Atrial Fibrillation Episodes from Sparse Electrocardiogram Data," *Journal of Electrocardiology,* vol. 41, 292-299, 2008.

[34] C. O'Brien and C. Heneghan, "A Comparison of Algorithms for Estimation of a Respiratory Signal from the Surface Electrocardiogram," *Computers in Biology and Medicine,* vol. 37, 305-314, 2007.

[35] H. Yang, *et al.*, "Nonlinear Adaptive Wavelet Analysis of Electrocardiogram Signals," *Physical Review E,* vol. 76, 026214, 2007.

[36] M. Richter and T. Schreiber, "Phase Space Embedding of Electrocardiograms," *Physical Review E,* vol. 58, 6392, 1998.

[37] F. Jager, *et al.*, "Long-Term St Database: A Reference for the Development and Evaluation of Automated Ischaemia Detectors and for the Study of the Dynamics of Myocardial Ischaemia," *Medical and Biological Engineering and Computing,* vol. 41, 172-182, 2003.

[38] J. D. Thomas, *et al.*, "Physical and Physiological Determinants of Pulmonary Venous Flow: Numerical Analysis," *American Journal of Physiology - Heart and Circulatory Physiology,* vol. 272, H2453-H2465, 1997.

[39] D. C. Chung, *et al.*, "A Dynamic Model of Ventricular Interaction and Pericardial Influence," *American Journal of Physiology Heart and Circulation Physiology,* vol. 272, H2942-H2962, June 1, 1997.

[40] J. T. Ottesen and M. Danielsen, "Modeling Ventricular Contraction with Heart Rate Changes," *Journal of Theoretical Biology,* vol. 222, 337-346, 2003.

[41] K. B. Campbell, *et al.*, "Validation of Optional Elastance-Resistance Left Ventricle Pump Models," *American Journal of Physiology - Heart and Circulatory Physiology,* vol. 251, H382-H397, 1986.

[42] K. Lu, *et al.*, "A Human Cardiopulmonary System Model Applied to the Analysis of the Valsalva Maneuver," *American Journal of Physiology Heart and Circulation Physiology,* vol. 281, H2661-H2679, December 1, 2001.

[43] R. M. Lewis and V. Torczon, "Pattern Search Algorithms for Bound Constrained Minimization," *SIAM Journal on Optimization,* vol. 9, 1082-1099, 1999.

[44] V. Torczon, "On the Convergence of Pattern Search Algorithms," *SIAM Journal on Optimization,* vol. 7, 1-25, 1997.

[45] M. R. Pinsky, *et al.*, "Changes in Electrocardiographic Morphology Reflect Instantaneous Changes in Left Ventricular Volume and Output in Cardiac Surgery Patients," *The American Journal of Cardiology,* vol. 76, 667-674, 1995.

[46] T. Feldman, *et al.*, "Change in Ventricular Cavity Size: Differential Effects on Qrs and T Wave Amplitude," *Circulation,* vol. 72, 495-501, September 1, 1985.

[47] S. Akselrod, *et al.*, "Power Spectrum Analysis of Heart Rate Fluctuation: A Quantitative Probe of Beat-to-Beat Cardiovascular Control," *Science,* vol. 213, 220-222, 1981.

[48] M. Pagani, *et al.*, "Power Spectral Analysis of Heart Rate and Arterial Pressure Variabilities as a Marker of Sympatho-Vagal Interaction in Man and Conscious Dog," *Circulation Research,* vol. 59, 178-93, August 1, 1986.

[49] A. C. Rencher, *Methods of Multivariate Analysis*, Second ed.: Wiley Series in Probabylity and Mathematical Statistics, 2002.

[50] J. Neaton, *et al.*, "Blood Pressure, Systolic and Diastolic, and Cardiovascular Risks: Us Population Data," *Archives of Internal Medicine,* vol. 153, 598-615, 1993.

[51] J. Stamler, *et al.*, "Blood Pressure (Systolic and Diastolic) and Risk of Fatal Coronary Heart Disease," *Hypertension,* vol. 13, I2-12, 1989.

[52] W. B. Kannel, *et al.*, "Systolic Versus Diastolic Blood Pressure and Risk of Coronary Heart Disease: The Framingham Study," *The American Journal of Cardiology,* vol. 27, 335-346, 1971.

[53] A. N. Pettitt, "A Two-Sample Anderson-Darling Rank Statistic," *Biometrika,* vol. 63, 161-168, 1976.

[54] F. W. Scholz and M. A. Stephens, "K-Sample Anderson-Darling Tests," *Journal of the American Statistical Assoication,* vol. 82, 1987.

[55] S. Konishi and G. Kitagawa, *Information Criteria and Statistical Modeling*: New York Springer, 2008.

[56] A. C. Guyton, *Textbook of Medical Physiology*, 8th ed.: Philadelphia: Saunders, 1991.

[57] G. A. Brecher and C. A. Hubay, "Pulmonary Blood Flow and Venous Return During Spontaneous Respiration," *Circulation Research,* vol. 3, 210-214, 1955.

[58] M. R. Pinsky, "Determinants of Pulmonary Arterial Flow Variation During Respiration," *Journal of Applied Physiology.,* vol. 56, 1237-1245, 1984.

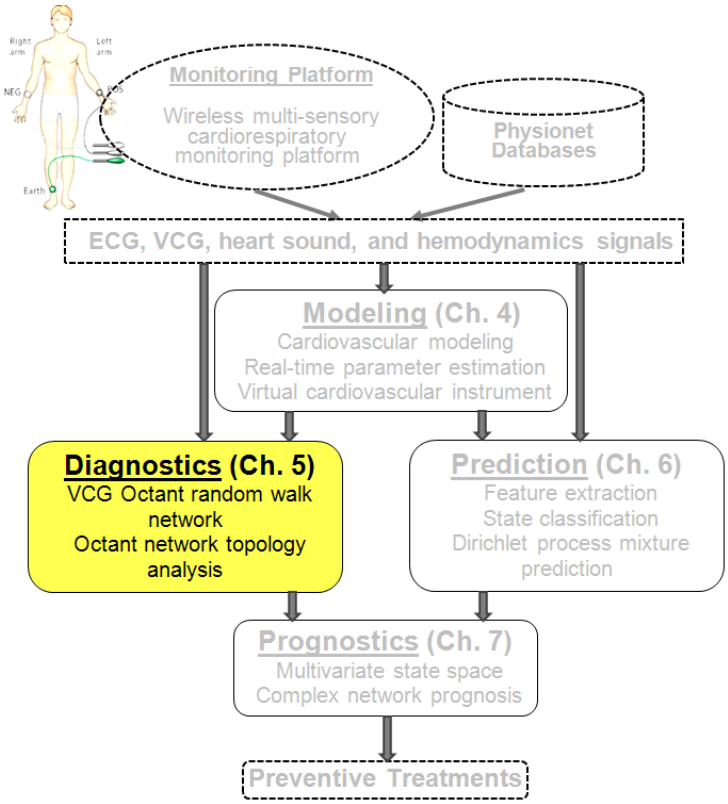
[59] I. B. Wilkinson, *et al.*, "The Influence of Heart Rate on Augmentation Index and Central Arterial Pressure in Humans," *Journal of Physiology,* vol. 525, 263-270, 2000.

[60] J. Smith, *et al.*, "Monitoring Arterial Blood Pressure and Cardiac Output Using Central or Peripheral Arterial Pressure Waveforms," in *Intensive Care Medicine*, ed Berlin: Springer, 2009, pp. 285-296.

[61] R. Pearse, *et al.*, "Equipment Review: An Appraisal of the Lidco Plus Method of Measuring Cardiac Output," *Critical Care,* vol. 8, 190 - 195, 2004.

CHAPTER 5

A RANDOM THEORETIC APPROACH FOR DISORDER DIAGNOSTIC AND LOCALIZATION



A high-specificity diagnostic approach systematically translates the measured and derived physiological signals into organized and classified conditions of the cardiovascular disorder. We present a diagnostic approach for the detection and localization of cardiovascular disorder based on representing complex spatiotemporal patterns of cardiac dynamics as a random walk network

reconstructed from the evolution of VCG signals across a 3-dimensional state space. Extensive tests with signals from PTB database of PhysioNet.org suggest that locations of MI can be determined accurately (sensitivity of ~88% and specificity of ~92%) from tracking certain consistently estimated invariants of this random walk representation. The high accuracy of the diagnostic approach presented in this chapter is critical for the state detection; hence, the prognostic performance of the prognostic model in chapter 7.

5.1 Introduction

Myocardial infarction (MI), also referred to as heart attack, is a major cause of mortality in the US and worldwide. Roughly half a million Americans have had a recurrent MI, and an estimated 800,000 new MI patients has been identified in 2012 [1]. MI results from a prolonged coronary artery occlusion and insufficient blood perfusion of the heart muscle (myocardium). When the blood supply to the myocardium is interrupted, the cardiac cells (myocytes) become dysfunctional or necrotic due to the lack of oxygen, causing a disruption in the regulation and flow of charged ions within the intracardiac conduction pathway [2]. Consequently, the complex dynamics that determines the instantaneous mean electric axis of the heart during the myocardial depolarization (QRS-loop in Figure 5-1 (b)) and repolarization (T-loop in Figure 5-1 (b)) is dramatically affected [3]. For example, during the depolarization, as the excitatory current (or wave) moves through the ventricular myocardium, nonconductive necrotic myocardium is bypassed. The deviation of the depolarization pathway from the normal trajectory after the MI events generates instantaneous changes specific to the location and size of the infarcted areas [2].

5.2 Background and literature review

The MI locations may be broadly classified into anterior, inferior, posterior, inferolateral, anteroseptal, anterolateral and posterolateral. Expeditious detection of the location, not just the

presence and extent of MI can dramatically increase survival rates by allowing customized procedures (e.g., reperfusion therapy, cardiac catheterization, or coronary bypass surgery) necessary to treat each type of MI [4]. Attempts have been made to use morphological features from electrocardiogram (ECG) signals for MI localization. For example, ST elevation and Q wave elongation in precordial leads V1-V3 suggest an acute anterior MI, ST elevation and Q-wave elongation in leads II, III, and AVF, along with ST depression in leads I and AVL suggest an inferior MI [5]. Similarly, ST elevation and Q elongation in leads I, AVL, V5, V6, together with ST depression in leads II, III, and AVF suggests a lateral MI [6]. In addition, features derived from wavelet transforms of ECG have also been attempted to detect MI location [7]. The limited success of these approaches have deemed the ECG alone insufficient, and concomitant elevated cardiac biomarkers to be necessary for MI localization [8]. Given the insensitivity of traditional ECG measures and the length of time before cardiac enzymes can be detected (> 5hrs), this recommendation underscores the need for features that can track the local spatiotemporal variations discernible from ECG to advance MI localization.

One technique that has garnered significant attention in recent years to visualize spatiotemporal distribution of electrocardiac activity is the vectorcardiogram (VCG) [9, 10]. The VCG, collected using Frank Vx, Vy, and Vz leads (see Figure 5-1 (a)) or transformed from the traditional 12-lead ECG, can capture much of the spatiotemporal behavior of instantaneous electrical potentials throughout each cardiac cycle in a 3-dimensional state space [11]. Numerous studies have attempted to quantify the spatiotemporal patterns, including the shape and magnitude of the P, QRS and T loops (Figure 5-1 (b)) in an effort to distinguish healthy individuals from patients with MI [12, 13]. While these attempts have been met with varied success, Laufberger's pioneering work introduced the possibility of using octant-wise (see Figure 5-2 (a)) distribution of VCG to more sensitively detect MI [14]. Due to computer processing limitations this potential

was not fully realized until the authors [10] extracted features that quantify the morphology of VCG in different octants to sensitively classify MI.

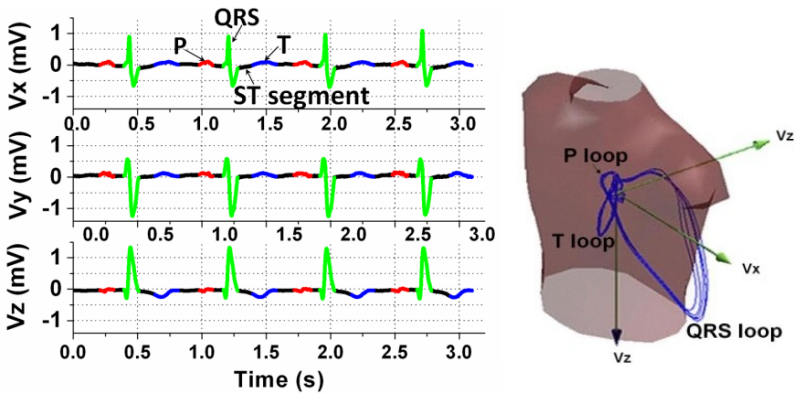


Figure 5-1 (a) Time-portrait of 3 VCG channels (Vx, Vy, and Vz) and (b) representative VCG trajectory showing P, T, and QRS loops embedded in a human torso-referenced Cartesian coordinate system [9]

However, the aforementioned features are not adequate for accurate localization of MI. This is because as mentioned in the foregoing, when VCG pathways change due to a specific MI, the sequence of transitions between octants, not just the VCG morphology varies at least locally. Consistent quantification of these local transitions is challenging because significant variations exist in how these transitions occur from beat to beat, even within a 20 sec long recording. These variations result from the effects of intrinsic and extraneous control mechanisms that determine or be determined by cardiac excitation and impulse propagation patterns through cardiac myocytes, as well as the effects of cardiorespiratory interactions [15, 16]. A probabilistic approach is therefore necessary to quantify these local transitions.

5.3 Research approach

5.3.1 Octant network representation

In the present investigation we represent spatiotemportal dynamics underlying the measured VCG signals as a near-stationary random walk network [10]. Random walk representations are

emerging as convenient means to analyze complex systems innate to physical, biological and social sciences, notably for extracting quantifiers of important network topology and dynamics, such as the salient nodes in a hyperlink network page-ranking algorithm, or convergence rate to a stationary distribution in sampling problems. Here, we identify a set of features that quantify the local topology of VCG random walk network (e.g., clustering, density, assortativity, centrality, transition rates, sojourn time and mixing rate) for sensitive estimation of MI locations, present a method to determine the length of a signal segment necessary to provide consistent estimate of these features, and discuss the physiological basis for their sensitivity.

The evolution of VCG across eight octants over a short recording (< 1 min) is represented as a random walk taken on a complete octant network $G(V)$, where nodes $V = \{1,2,\dots,8\}$ represent the different octants (Figure 5-2 (a)). This random walk process essentially captures the spatial transition dynamics, as discerned from VCG, of the cardiac vector in octant space during the depolarization and repolarization processes of a cardiac cycle, and it can be represented as a directed random graph $G(V, E, A)$ (see Figure 5-2 (b)). Here the adjacency matrix A is the transition matrix of the Markov chain underlying the random walk, i.e., $a_{ij} = p_{ij}$, the transition probability between nodes (octants) i and j, and the edge set $E = \{e_1, e_2, \dots, e_l\}$ is the collection of the plausible transitions i.e., $p_{ij} > 0$. It may be noted that the adjacency matrix is nonnegative, generally asymmetric, and has zero diagonal (since transitions happen only when a VCG vector crosses an octant boundary). Thus, the aperiodic and recurrent patterns in a VCG trajectory (which may be viewed as an output of cardiac processes) are represented in terms of an irreducible Markov chain formed by the random walk on the octant network (Figure 5-2 (c)), and spatiotemporal cardiac dynamics underlying VCG can be captured by estimating the random walk quantifiers and topological measures of octant transition network.

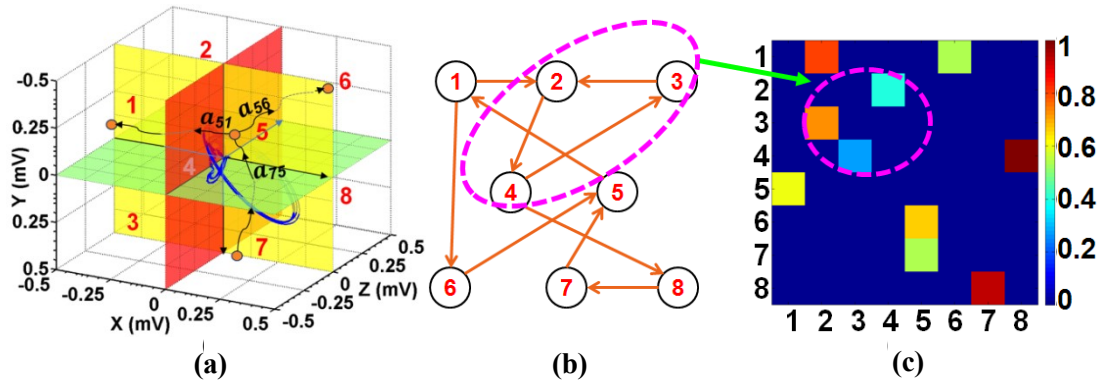


Figure 5-2 (a) Representation of VCG transitions between octants as a random walk, where a_{ij} is the transition probability from octant i to octant j and a_{ji} from octant j to octant i. Generally $a_{ij} \neq a_{ij}$ leading to a directed weighted graph representation of VCG random walk; (b) A directed weighted graph representation of stochastic transitions of VCG trajectory in the octant space; (c) Normalized and color-coded adjacency matrix of the undirected weighted graph from (b) where the transition probabilities are scaled such that the largest probabilities in the graph equals 1 (deep red color). Complex network measures extracted from this adjacency matrix are used for localizing MI.

5.3.2 Random walk on the octant network

The chief quantifiers of a random walk may be expressed in terms of a fundamental matrix [17] $Z = \sum_{t=0}^{\infty}(P^t - J\Pi) = (I - P + J\Pi)^{-1} - J\Pi$, where $\Pi = diag(\pi_i)$ is the diagonal matrix containing the stationary probabilities π_i's, P is the transition probability matrix, and the matrix $J = [J_{ij}]$ is such that $J_{ij} = 1$, $1 \leq i, j \leq 8$. In this context, as random walk proceeds from an arbitrary initial node marked by the starting point of a VCG segment, the probability $\pi_i(n)$ of finding the walk at a node i becomes independent of the number n of transitions. The rate μ at which such a (stationary) random walk converges to this limiting distribution π_i is called the mixing rate. Mixing rate can be used to estimate the convergence rate of the measures extracted from cardiac dynamics reconstructed from VCG to those of a stationary attractor.

The mixing rate of a random walk can be expressed in terms of a normalized digraph Laplacian $\tilde{\mathcal{L}} = \Pi^{\frac{1}{2}}(I - P)\Pi^{-\frac{1}{2}}$ [18]. To account for the asymmetry of a directed weighted graph representing the VCG random walk, $\tilde{\mathcal{L}}$ is decomposed as a sum of a symmetric and a skew-

symmetric part, i.e., $\tilde{\mathcal{L}} = \bar{\mathcal{L}} + \Delta$ where $\bar{\mathcal{L}} = (\tilde{\mathcal{L}} + \tilde{\mathcal{L}}^T)/2$ and $\Delta = (\tilde{\mathcal{L}} - \tilde{\mathcal{L}}^T)/2$. Let $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_8$ be the singular values of $\tilde{\mathcal{L}}$, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_8$ the eigenvalues of $\bar{\mathcal{L}}$, and $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_8$ the singular values of Δ. The following result holds on the mixing rate of the random walk on the octant transition network [18]:

Result 1. *The mixing rate μ of the random walk on the octant transition network remains bounded by the singular values as*

$$\delta_8^2 \leq \mu \leq (1 - \lambda_2)^2 + 2\delta_8\lambda_2 + \delta_8^2$$

It may be noted that maximum likelihood estimates $\hat{P}(n)$ updated under stationary conditions from observations after every transition n converge in mean to P [19]. Consequently, the estimates of the singular values λ_k and δ_k converge, and hence the mixing rate estimates $\hat{\mu}(n)$ converge in mean to μ.

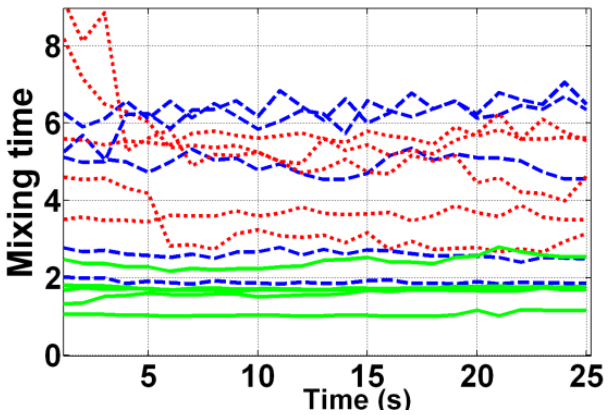


Figure 5-3 Variation of mixing time of VCG octant random walk network for recordings from inferior MI (dashed blue line), anterior MI (dotted red line), and healthy control subjects (solid green line). The signal segments were 25 sec long. It may be noted that the mixing times converge to their limiting values in three cases after 10 sec

The mixing time μ^{-1}, defined as the expected number of transitions needed for a random walk process to converge to a distribution, can be used to estimate the number of transitions, and

hence the signal length necessary to provide a consistent estimate of the random walk quantifiers. As shown in Figure 5-3, the mixing rates of VCG random walk estimated based on $\hat{P}(n)$ stabilize after about 8 sec, corresponding to roughly 12 beats. Overwhelmingly, the mixing times converge to values < 3 transitions for healthy control (HC) recordings indicating fairly laminar VCG trajectories. However for MI recordings, mixing rates converge to between 2.5-7 transitions, indicating an increase in erratic movements of VCG trajectories. Based on this result, we have therefore used 10 sec long signal segments to derive consistent estimates of various features for MI localization (see 5.4).

Additionally, hitting time [17] $H_{ij} = (z_{jj} - z_{ij})/\pi_j$ and commute time $C_{ij} = \left[(z_{jj} - z_{ij})/\pi_j + (z_{ii} - z_{ji})/\pi_i\right]$ of a random walk can be used to estimate the expected lengths of the pathways between any two octants i and j, as well as those of various VCG loops. A notion of length or distance signifies the pathway of propagating electric waves along the cardiac conduction system, and is therefore important for defining topological measures for the VCG octant transition network. As stated in the foregoing, manifestations of MI modify the conduction pathways, especially near infarcted areas, thereby altering the path-lengths of the waves. The local topology measures of the octant transition network quantify these changes in the transition rates and VCG trajectory patterns. For example, Figure 5-4 shows the transition matrix P estimated for 6 different groups in the database, namely healthy control (HC), inferior (I), inferolateral (IL), anterior (A), anteroseptal (AS), anterolateral (AL). We group I and IL into inferior MI family (IF) and A, AS and AL into anterior MI family (AF). There exists a significant difference between the transition matrices estimated for different groups. In particular, the estimated value of $\sum_{i=1}^{4} P_{ij}$ was much higher for inferior MI compared to those estimated for anterior MI. This is reasonable because trajectories of VCG cardiac vector within octants 1 to 4

(the right side of the sagittal plane) are the measurements of the action potential propagating on the surface of cardiac muscle in the inferior portion of the heart. Inferior necrosis damages the heart muscles in these octants. As a result VCG trajectory meanders away from these octants thereby reducing the transitions across these octants. Thus the random walk network topology can be spatially associated with the cardiac conduction process.

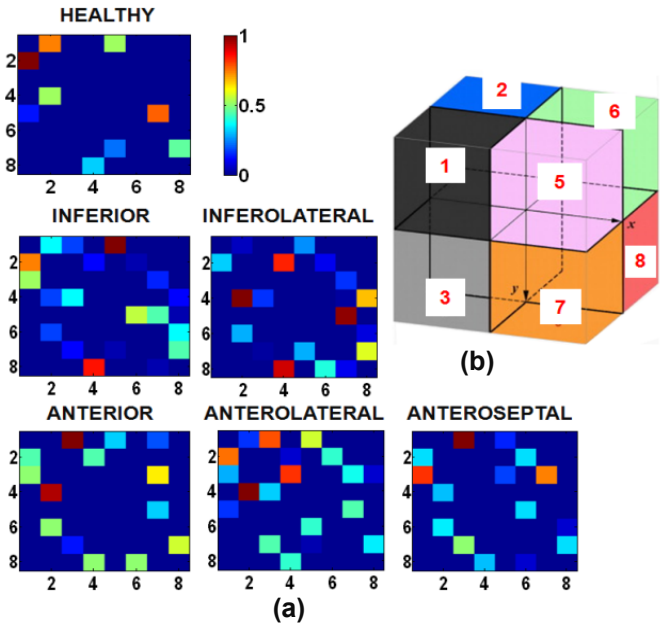


Figure 5-4 a) Transition probabilities of VCG random walk for three different groups of patients (color coded based on the transition probability estimates) b) locations of eight octants (1-8) in a 3-D space [10]

Based on the foregoing qualitative observations of how the characteristics of the VCG and the topology random walk network constructed therefrom vary with MI manifestation, we had extracted following four groups of 161 features from the measured signals and their random walk representation for MI localization:

(1) Local octant features: These include 48 features, such as the minimum, average, variance, maximum, azimuth, and elevation of the vector magnitudes. These features have been employed in our earlier study for MI identification [10].

(2) Octant residence features: The 12 features in this group include sojourn times in each octant, and the velocity of the vector magnitudes.

(3) Octant transition features: The16 features in this group include the arrival and the departure rates between each pair of octants. These are essentially the estimates of elements p_{ij} of the transition matrix *P*.

(4) Network topology features: The 85 of this group are various quantifiers of the topology of the random walk network, including the degree assortativity (48), density clustering (17), distances and cycles that include those related to the estimates of H_{ij} and C_{ij} (12), and betweenness centrality (8) [20].

These features were used as part of a hierarchy of classification and regression tree (CART) models [21] described in the following section to identify the various MI locations.

5.4 Implementation details and results

The method was tested using VCG signals from PTB database of Physionet [22]. Some 448 VCG recordings (368 MIs and 80 HCs) available in the PhysioNet PTB Database were considered for this investigation. Each recording contains 15 simultaneous heart monitoring signals, namely, the conventional 12-lead ECG and the 3-lead Frank (XYZ) VCG. The signals were digitized at 1 kHz sampling rate with a 16-bit resolution over a range of ±16.384 mV. The 80 recordings were acquired from 54 healthy subjects and the 368 recordings from 148 MI patients. 282 acute MI recordings without former MI were investigated in this study. Among the acute MI recordings, 117 belonged to inferior family (IF) with 74 inferior (I) and 43 inferolateral (IL), 133 to anterior family (AF) with 35 anterior (A), 42 anterolateral (AL) and 56 anteroseptal (AS), and 32 with less common or undetermined locations. The VCG signals were recorded for at least 30 sec with an average length of ~2 min.

We employed a classification and regression tree (CART) –an established method for creating clinical rules for medical diagnosis [21]– to distinguish between the five different MIs and HC groups using the aforementioned features. A CART model recursively partitions the multidimensional space formed by 161 features (48 octant morphology features [10], 16 P_{ij} values, 85 network topology features [20] and 12 octant sojourn time features) into subsets of the feature space, each representing the feature values for HC or each type of MI. Such high-dimensional feature space containing highly redundant information and with limited number of recordings degrades the classification accuracy and efficiency. Anderson-Darling (AD) and Kolmogorov Smirnov (KS) tests were therefore applied to yield a subset of 10 features (see Table 5-1) that are significant for MI localization.

Table 5-1 Summary of the selected features employed in the optimal classifiers (CART 1, CART 2, and CART 3)

CART MODEL	FEATURE DEFINITION	DESCRIPTION
CART 1 Inferior family Anterior family Healthy control	Oct7Var = var($\mathbf{V}_7$) where $\mathbf{V}_7$ is VCG vector magnitude set in octant 7	Vector magnitude variance in octant 7
	InStr6 = $\sum_{j=1,j\neq 6}^{8} w_{j6}$ where w_{j6} is the inward link weight from octant j to octant 6	Sum of inward link weights to octant 6
	Oct1Var = var($\mathbf{V}_1$) where $\mathbf{V}_1$ is VCG vector magnitude set in octant 1	Vector magnitude variance in octant 1
	Oct8Var = var($\mathbf{V}_8$) where $\mathbf{V}_8$ is VCG vector magnitude set in octant 8	Vector magnitude variance in octant 8
	Oct2Num = $\sum_{k=1}^{n} T_k^2$ where T_k^2 sojourn time in octant 2 of VCG trajectory in heart cycle k^{th}, and n is the number of heart cycles	Summation of sojourn time in octant 2 of the VCG vector
CART 2 Inferior Inferolateral	Clust2 = $\frac{t_2}{(k_2^{out}+k_2^{in})\left(k_2^{out}+k_2^{in}-1\right)-2\sum_{j\in N} a_{2j}a_{j2}}$ where $t_2$is the number of triangles around octant 2; k_2^{out} and k_2^{in} are the number of outward and inward links from octant 2; and a_{ij} is the connection status between *i* and *j*: a_{ij} = 1 when link (i, j) exists (when i and j are neighbors); a_{ij}= 0 otherwise (a_{ii} = 0 for all *i*)	Average number of directed-triangles around octant 2
	Oct4Max = max($\mathbf{V}_4$) where $\mathbf{V}_4$ is VCG vector magnitude set in octant 4	Amplitude of the maximal vector in octant 4
CART 3 Anterior Anterolateral Anteroseptal	OutOct4Rate (μ_4) = $\frac{\sum_{j=1}^{8}\lvert p_{4j}\rvert}{\sum_{i}^{8}\sum_{j=1}^{8}\lvert p_{ij}\rvert}$ where p_{ij} is the expected probability of transition from octant i to octant j updated from the observations under stationary condition	Departure rate of VCG trajectory from octant 4 to all other octants

	Clust5 = $\frac{t_5}{(k_5^{out}+k_5^{in})(k_5^{out}+k_5^{in}-1)-2\sum_{j\in N} a_{5j}a_{j5}}$ where $t_5$is the number of triangles around octant 2; k_5^{out} and k_5^{in} are the number of outward and inward links from octant 5; and a_{ij} is the connection status between *i* and *j*: a_{ij} = 1 when link (i, j) exists (when i and j are neighbors); a_{ij}= 0 otherwise (a_{ii} = 0 for all *i*)	Average number of directed-triangles around octant 5
	InDrg3 (k_3^{in}) = $\sum_{j\in N} a_{j3}$ where *N* is the number of octants in the network	Number of inward transitions connected to octant *3* from all other octants

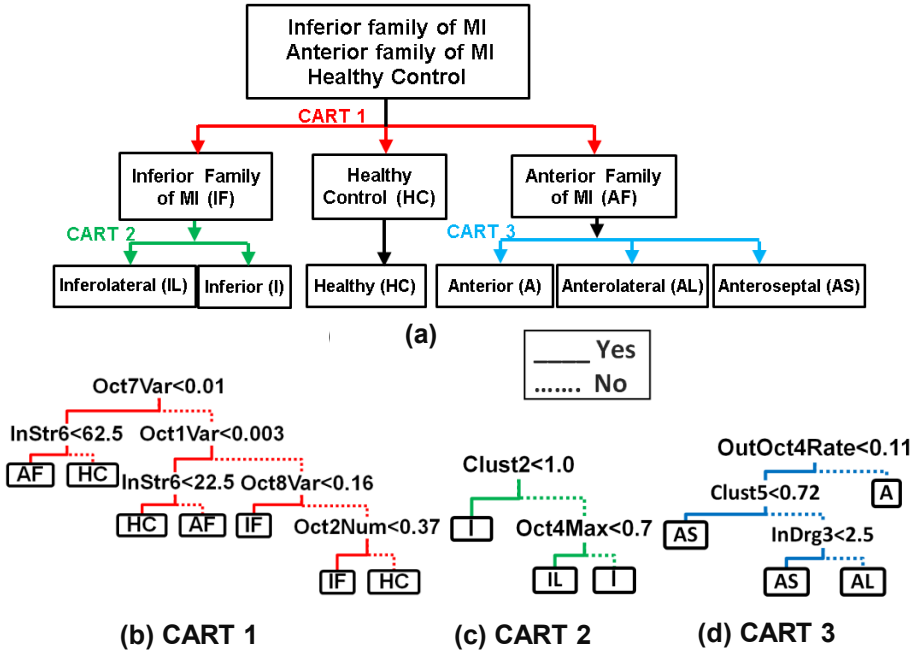


Figure 5-5 a) A summary of the hierarchical CART models to classify Inferior (IF) and Anterior (AF) MIs from Healthy Control (CART 1); two subgroups of inferior MIs, namely, Inferolateral (IL) and Inferior (I) (CART 2); and three subgroups of anterior MIs, namely, Anterior (A), Anterolateral (AL), and Anteroseptal (AS) (CART 3). All CART models are specified in terms of a treestructure with the solid lines denoting the TRUE branch (i.e., the condition stated at a node of the tree holds) and the dashed line denoting the FALSE branch. The optimized CART 1, CART 2, and CART 3 model structures are showed in (a), (b), and (c), respectively.

While morphological features of VCG octants were largely adequate to classify between HC and MI, transition features were found to be highly sensitive to MI location (see Figure 5-5). Here, CART 1 (Figure 5-5 (b)) provides the rules to distinguish between HC, IF and AF, CART 2 (Figure 5-5 (c)) between I and IL, and CART 3 (Figure 5-5 (d)) between A, AS and AL. Evidently, the transition probabilities (e.g., OutOct4Rate) and topology quantifiers (e.g, InDrg3,

Clust2, Clust5) of the VCG random walk network are the key features for identifying the various MI locations. The results from the hierarchical classifications, summarized in Table 5-2, suggest that the accuracies of classification among 5 types of MIs and HC exceed 85% with an average sensitivity of ~88% and average specificity ~92%. The results also indicate that compared to AF cases, the IF cases (inferior and inferolateral MI) can be more accurately localized. The lower classification accuracies for the AF family may be attributed to the spatial overlap of the affected regions in the three anterior MI groups.

Table 5-2 Summary of the results from hierarchical CART classification: Normalized confusion matrix from (a) CART 1, (b) CART 2, and (c) CART 3 classifiers suggest that MI locations can be identified with sensitivity of > 84%, and specificity of > 86%. (d) Classification accuracies calculated from the confusion matrices suggest that the accuracies of all classification cases exceed 85%. Inferior MI family (I and IL) can be identified most accurately with the best classification results in terms of high (> 90%) sensitivity and specificity.

Act. / Class'n	IF	A F	HC
IF	0.94	0.03	0.06
AF	0.02	0.93	0.07
HC	0.04	0.04	0.87

(a)

Act. / Class'n	I	IL
I	0.9	0.09
IL	0.1	0.91

(b)

Act. / Class'n	A	AS	AL
A	0.88	0.02	0.02
AS	0.06	0.85	0.14
AL	0.06	0.13	0.84

(c)

Groups / Metric	IF	AF	HC	I	IL	A	AS	AL
Accuracy	0.91	0.91	0.91	0.91	0.91	0.86	0.86	0.86

(d)

5.5 Discussion

Thus, complex network analysis is used to capture the relationship between local variations in physiological conduction system of the heart and the spatiotemporal dynamics of VCG. The results suggest that the network measures identified in the present work can be used to characterize the propagation patterns of the cardiac action potential during the depolarization (roughly contemporaneous with heart contraction and manifestation of P and QRS loops in VCG)

and repolarization (contemporaneous with heart relaxation and manifestation of T loop) phases for accurate MI localization. Parenthetically, an atrial depolarization phase is marked by the generation of rhythmical impulse from the sinus node and its propagation through the atrioventricular node via the intermodal pathways, and ventricular depolarization by the spread of impulse signal through A-V bundle, Purkinje fibers, and to all parts of the ventricles. The spread of impulse for repolarization of the heart for the next contraction is marked by a T wave. The cardiac conduction patterns are susceptible to the damaged heart tissues resulting from MI. Accurate MI localization from using the topological and dynamic features of the random walk network support the clinical evidences that the distortion of the contour and orientation of QRS and T loops in VCG signals provide an accurate diagnosis for MI [23].

5.6 Conclusions

In summary, we have introduced a method for MI localization based on capturing the complex cardiac excitation and propagation dynamics as a random walk network reconstructed from VCG signals. Various topological and dynamic quantifiers of the random walk network were found to be sensitive to the location of cardiac tissues damaged as a result of an MI and, consequently, they serve as effective features for MI localization. Extensive tests conducted using data from PTB database of PhysioNet suggest that hierarchical CART classifier to identify different MI types has a sensitivity of ~88% and specificity of ~92% (standard deviation <5%). Such random walk network quantifiers can be applied to detect various cardiovascular disorders related to the deviation of the intracardiac conduction pathways such as tachycardia and bradycardia. It may be noted that the VCG signals wherefrom the octant network is reconstructed can be derived from 12-lead ECG signals using Dower and other transformations [24]. Consequently the present method can be implemented as part of a diagnostic system for MI localization based on ECG signals (3 or 12 lead) recorded from standard ECG machines, which

are accessible to most medical facilities. The ongoing investigations are aimed at capturing more fine-grained transitions within an octant (beyond the currently used transitions between octants) as well as deriving separate network representations for various cardiac processes (repolarization versus depolarization) towards further enhancing the sensitivity and specificity of MI localization.

REFERENCES

[1] V. L. Roger, *et al.*, "Heart Disease and Stroke Statistics—2012 Update," *Circulation,* vol. 125, e2-e220, January 3, 2012.

[2] L. H. Opie, *Heart Physiology: From Cell to Circulation*, 3rd ed.: Lippincott-Raven Philadelphia, 1998.

[3] W. Carson, *et al.*, "Vectorcardiographic Criteria for Acute Right Ventricular Infarction," *European Heart Journal,* vol. 9, 955-961, September 1, 1988.

[4] S. E. Epstein, *et al.*, "Evaluation of Patients after Acute Myocardial Infarction," *New England Journal of Medicine,* vol. 307, 1487-1492, 1982.

[5] R. A. Warner, *et al.*, "Electrocardiographs Criteria for the Diagnosis of Anterior Myocardial Infarction: Importance of the Duration of Precordial R Waves," *The American Journal of Cardiology,* vol. 52, 690-692, 1983.

[6] G. B. Myers, *et al.*, "Vii. Correlation of Electrocardiographic and Pathologic Findings in Lateral Infarction," *American Heart Journal,* vol. 37, 374-417, 1949.

[7] A. N. Neskovic, *et al.*, "Myocardial Tissue Characterization after Acute Myocardial Infarction with Wavelet Image Decomposition," *Circulation,* vol. 98, 634-641, August 18, 1998.

[8] K. Thygesen, *et al.*, "Universal Definition of Myocardial Infarction," *European Heart Journal,* vol. 28, 2525-2538, October 1 2007.

[9] H. Yang, *et al.*, "Spatiotemporal Representation of Cardiac Vectorcardiogram (VCG) Signals," *BioMedical Engineering OnLine,* vol. 11, 16, 2012.

[10] H. Yang, *et al.*, "Identification of Myocardial Infarction (Mi) Using Spatio-Temporal Heart Dynamics," *Medical Engineering and Physics,* vol. 34, 485-497, 2012.

[11] L. Wolff, *et al.*, "Spatial Vectorcardiography," *New England Journal of Medicine,* vol. 248, 810-819, 1953.

[12] J. W. Starr, *et al.*, "Vectorcardiographic Criteria for the Diagnosis of Inferior Myocardial Infarction," *Circulation,* vol. 49, 829-836, May 1, 1974.

[13] L. Edenbrand, *et al.*, "Vectorcardiogram More Sensitive Than 12-Lead Ecg in the Detection of Inferior Myocardial Infarction," *Clinical Physiology,* vol. 10, 551-559, 1990.

[14] V. Laufberger, "Octant Vectorcardiography," *Physiol Bohemoslov,* vol. 29, 481-94, 1980.

[15] D. R. Chialvo, *et al.*, "Low Dimensional Chaos in Cardiac Tissue," *Nature,* vol. 343, 653-657, 1990.

[16] V. N. Smelyanskiy, *et al.*, "Inference of a Nonlinear Stochastic Model of the Cardiorespiratory Interaction," *Physical Review Letters,* vol. 94, 098101, 2005.

[17] D. Aldous and J. A. Fill, "Reversible Markov Chains and Random Walks on Graphs," http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[18] Y. Li and Z.-L. Zhang, "Random Walks on Digraphs, the Generalized Digraph Laplacian and the Degree of Asymmetry Algorithms and Models for the Web-Graph." vol. 6516, ed: Springer Berlin / Heidelberg, 2010, pp. 74-85.

[19] L. Lovasz, "Random Walks on Graphs: A Survey," *Combinatorics,* vol. 2, 1993.

[20] S. Boccaletti, *et al.*, "Complex Networks: Structure and Dynamics," *Physics Reports,* vol. 424, 175-308, 2006.

[21] L. Goldman, *et al.*, "A Computer Protocol to Predict Myocardial Infarction in Emergency Department Patients with Chest Pain," *New England Journal of Medicine,* vol. 318, 797-803, 1988.

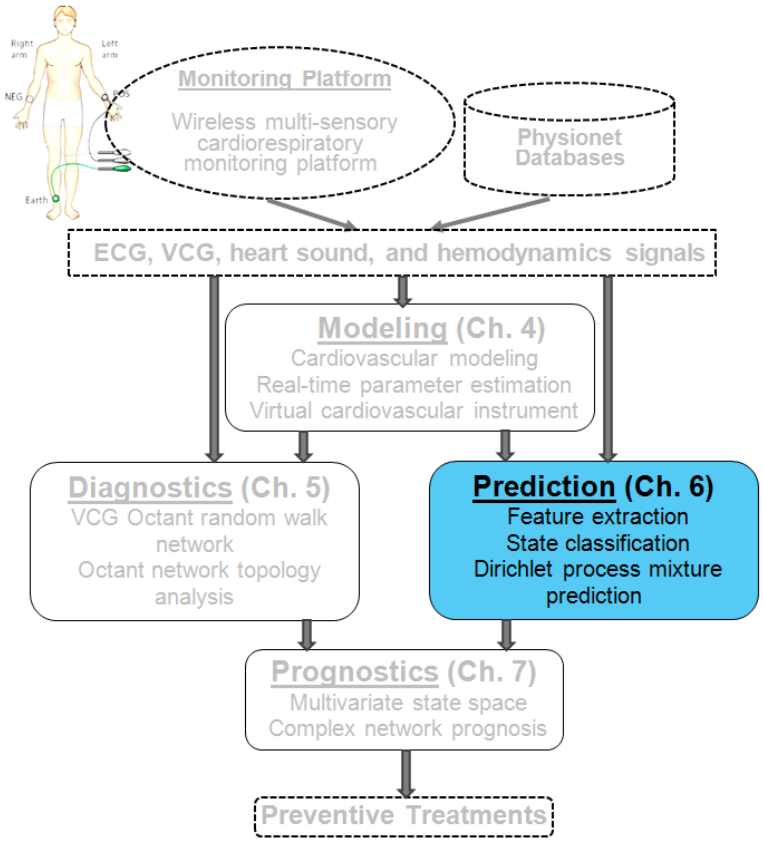
[22] A. L. Goldberger, *et al.*, "Physiobank, Physiotoolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation,* vol. 101, E215-20, Jun 13 2000.

[23] A. R. Pérez Riera, *et al.*, "Significance of Vectorcardiogram in the Cardiological Diagnosis of the 21st Century," *Clinical Cardiology,* vol. 30, 319-323, 2007.

[24] D. Dawson, *et al.*, "Linear Affine Transformations between 3-Lead (Frank Xyz Leads) Vectorcardiogram and 12-Lead Electrocardiogram Signals," *Journal of Electrocardiology,* vol. 42, 622-630, 2009.

CHAPTER 6

DIRICHLET PROCESS BASED MIXTURE GAUSSIAN PROCESS MODELS FOR PREDICTION OF DISORDER EVOLUTION



This chapter describes a prediction method to forecast the future states of the disorder evolution using the measured and derived physiological signals. The distribution of the predicted values provides a comprehensive risk indicator metric for the performance of the prognostic model investigated in chapter 7. We introduce a Dirichlet process-based mixture Gaussian

process (DPMG) model to predict the onset of abnormal state based on tracking complex evolution of the cardiovascular dynamic underlying the extracted signatures. Extensive testing with signals from the multisensory suite developed at COMMSENS lab and PhysioNet's OSA to forecast the onset of obstructive sleep apnea (OSA) suggests that the accuracies (R^2) are 83% and 77% for 1-min and 3-min ahead prediction, respectively.

6.1 Introduction

Obstructive sleep apnea (OSA) is a common sleep disorder that affects 24% of adult men and 9% of adult women [1]. An OSA episode is marked by the obstruction of pharyngeal airways and interruption of the airflow during sleep. It is known to reduce the sleep quality and other allied physiological processes vital for the cognitive and restorative functions [2]. Furthermore, due to the irregular sympathetic stimulation at the end of obstructive phase untreated OSA patients are also at increased risk for developing cardiovascular disorders, such as hypertension, coronary artery diseases, and stroke [3].

Noninvasive ventilation therapies, such as continuous positive airway pressure (CPAP) that delivers air with a pre-determined fixed pressure into the pharynx, and other oral appliances designed to prevent throat constrictions and deliver oxygen to the lung continuously, are used to treat OSA. Such treatments can minimize the sleep apnea (measured in terms of a apnea-hypoapnea index AHI) and thereby improve sleep quality, subjective wellness, and mental health, thus reducing stroke risk [4].

Despite its wide use, an estimated 46-83% of patients with OSA have been reported to be nonadherent to the use of CPAP beyond 4 hours. This is because CPAP devices often over-treat air supply to accommodate the differences in respiration patterns during the different sleep stages (e.g., REM, N1, N2) [5], and such excessive airflow oftentimes leads to extreme dryness of nose

and throat [4]. Real-time monitoring of OSA can be vital towards calibrating airflow to address this issue. Even better, the prediction of OSA events would allow proactive adjustment of airflow and body positions to mitigate OSA and improve the adherence of the patient to the CPAP therapy. Such a prediction-based approach needs a wearable multisensory suite for continuous data acquisition during sleep and methods to track and forecast the evolution of cardiorespiratory dynamics from measured signals. While some advances in wearable sensors for sleep monitoring have been reported, few, if any work has been reported towards using signals acquired from these units for prediction of OSA episodes.

In this paper, we introduce a method based on using data gathered from a wireless wearable multisensory suite to predict the occurrence of sleep apnea events. A unique wireless wearable multisensory suite is developed to continuously collect the cardiac and respiratory signals in real-time during sleep. Quantifiers of the coupled nonlinear and nonstationary cardiorespiratory dynamics underlying the measured signals are used as the inputs to predict the onset of sleep apnea events. We developed a novel DPGM model to predict the complex evolution of the OSA signatures. As mentioned in the foregoing, these predictions can be timely for automatic adjustment of airway pressure patterns in CPAP, supplemental oxygen devices, or to change the body posture such as slight adjustment of torso or chin position that avert the collapse of the airways, which precedes an OSA episode. Continuous OSA monitoring and prediction using the wireless wearable multisensory suite can thus improve the efficiency of OSA treatment, especially in out-of- hospital, at-home conditions.

6.2 Background and literature review

The field of medicine is on the verge of transformation where healthcare would be provided on a personal basis to prevent the illness rather than treat it post-trauma. This systems approach to

personalized healthcare is based on integrating concepts of systems biology and medicine known as (P4) personalized, predictive, preventive and participatory medicine [6]. Much of the current P4 emphasis is on collecting physiological data from ECG, CAT scan, genomic data, diet, etc., into large data warehouses and using advanced information infrastructures for predicting and monitoring chronic non-communicable diseases [6, 7]. It was also noted that early detection of acute disease episodes through noninvasive monitoring is effective for patients with chronic disorders because treatment costs escalate exponentially with delay in detection [8].

Among the chronic conditions, OSA and sleep-related breathing disorders affect the fourth of the US population [9]. Several OSA detection and prediction approaches based on correlating the statistical patterns of heart rate, respiration rate, and oxygen saturation (SpO2) signals during OSA episodes have been attempted [10]. For example, spectral energy of intrinsic mode functions were extracted from empirical mode decomposition of flow rate signals (from a CPAP machine) to estimate likelihood of OSA episodes [11]. Similarly, support vector machines (SVMs) developed using linear, polynomial and radial basis kernel functions, networks, clustering algorithms with wavelet features have been applied to distinguish cases with OSA from those which do not have sleep apnea [12]. Although, considerable attention of OSA detection methods has been given, prediction of (forecast) an impending OSA episode, necessary for calibrating CPAP therapy, have not been reported in literature. The few current reported (e.g., dynamic belief networks [13, 14]) use limited data from OSA patients to predict OSA episodes ~ 1 sec ahead or just predict the evolution of the physiological signals (i.e., heart rate, chest volume, blood oxygen saturation). These methods do not capture variations in nonlinear and nonstationary dynamics of the cardiorespiratory system responsible for the onset of OSA or sleep-related breathing disorder events. Also, the development of a wearable multisensory unit that would facilitate gathering of

signals necessary of prediction without causing palpable discomfiture remains elusive. The present work is aimed at addressing these gaps.

The present approach uses a wearable multisensory wireless unit customizable to the specific conditions of the patient such as age, gender, BMI, and diseases for continuous monitoring and prediction of OSA episodes. Also, since underlying cardiorespiratory dynamics is nonlinear, signal analysis and prediction methods based on the nonlinear and nonstationary characteristics were investigated to establish these relationships.

6.3 Research approach

The key contributions of the present research are in (a) the development of an economical wearable wireless multisensory unit capable of measuring signals essential for sleep monitoring, including ECG, heart sound, respiration, and SPO2 synchronously without causing posing significant discomfort or constraints on motion, and (b) a method to provide accurate prediction of an impending OSA episode by considering the nonlinear and nonstationary cardiorespiratory dynamics underlying the measured signals and the features extracted therefrom. As summarized in Figure 6-1, we use the data from sleep apnea-ECG database as well as signals gathered from our wearable multisensory unit for training and testing of the predictor and classifier. While the PhysioNet database consists of signals gathered from chronic OSA patients, the signals from the wearable multisensory unit were gathered from healthy subjects (to assess false positive rates). Various quantifiers of topology of the nonlinear attractor of cardiorespiratory dynamics reconstructed from the measured signals, including laminarity, determinism, entropy, recurrence rate were extracted as features ϑ to identify an OSA event using a support vector machine (SVM) classifier. The evolution of $\vartheta(t)$ was tracked using a nonparametric Dirichlet process based Gaussian mixture (DPMG) prediction method that effectively captures nonlinear nonstatationary

evolution. The k-step (minutes) look-ahead predictions $\hat{\vartheta}(t+k)$ of feature values were used to detect an impending OSA episode 1-3 minutes earlier with an accuracy of 70-90%. Such predictions can be vital to initiate adjustments or therapeutic interventions to avert an impending OSA episode [15]. The remainder of this section describes the two main contributions of this work, namely the multisensory suite and the prediction method.

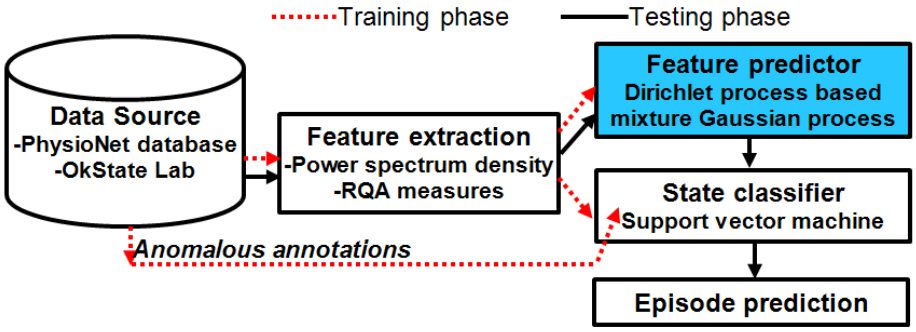


Figure 6-1 Overall approach for OSA episode prediction

6.3.1 Wireless wearable multisensory platform

We have developed a multisensory platform capable of synchronously gathering multiple heterogeneous signals, including VCG, ECG, sound, and respiration (see Figure 6-2 for screenshot of real-time streaming VCG, 3-D color coded VCG, and a standard display of 12-lead derived ECG), and wirelessly transmit the data to a host computer for on-line OSA prediction and subsequent therapeutic decision support. Such multi-channel data is necessary to track the dynamic decouplings known to precede the state transitions that lead to the onset of OSA episodes. Novel aspects of the proposed multi-sensor unit are as follows. (1) The sensors are judiciously chosen to capture the complementary aspects of the heart operation, *viz*. electrical (ECG), acoustic (sound), and mechanical (respiration). (2) Due to the use of MEMS technology, the total footprint of the wireless unit is highly adjustable and remains lightweight, and hence highly wearable. (3) The hardware platform in this context contributes towards affordable, yet

powerful, early warning (prognostic) systems for sleep apnea treatment. (4) The wireless (Bluetooth) platform along with the sensors and microprocessor components are integrated into a customized garment to continuously monitor and predict sleep apnea episodes [16]. The wireless design utilizes a class I Bluetooth device with response frequency range of 0.176-90Hz, sampling rate of up to 2 kHz, and 16 bit resolution.

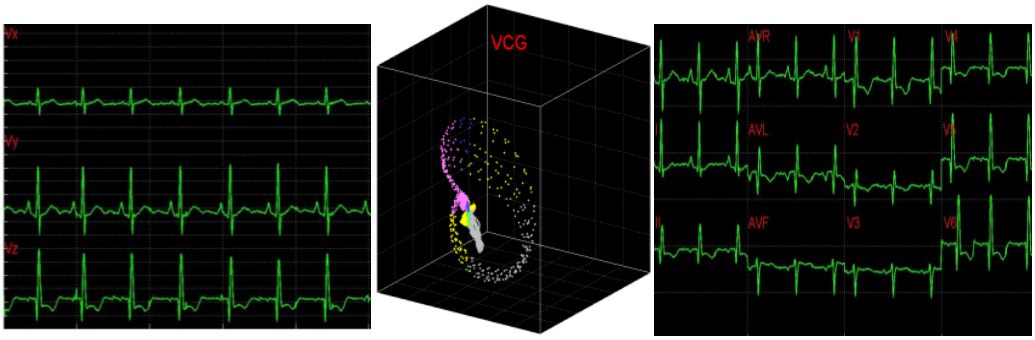


Figure 6-2 Screenshot of 3-channel streaming VCG, 3-D color coded dynamic VCG, and 12-lead transformed ECG signals

Figure 6-3 Wireless wearable multisensory suite

We have embedded the multisensory platform as part of a garment (see Figure 6-3) to enhance the wearability of the sensor suite. The fusion of information from VCG, heart sound, and respiration provide adequate information to track variations and detect transitions in cardiorespiratory dynamics during sleep.

6.3.2 Prediction model

Among nonparametric prediction models, Gaussian process (GP) can simplify the modeling efforts, but the computational overhead with covariance matrix inversion scales as a cube of the signal length and the assumption of a stationary covariance function impede its wide applications. In our previous work [17], we used local Gaussian process (LGP) to address the nonstationary issue. We used local topological characteristics including recurrence properties of dynamics reconstructed from a signal to partition the signal into near-stationary segments. While prior investigations suggest that LGP can be effective in predicting the evolution of nonlinear nonstationary processes, one needs to specify the threshold of a correlation index to determine the boundaries in the state space. Furthermore, estimation of the topological characteristics (in the form of recurrence maps) can be computationally intensive and somewhat sensitive to parameters such as the threshold. Towards addressing this limitation, we have investigated a Dirichlet process based Gaussian process mixture (DPMG) model to predict the evolution of the signal features.

In the DPMG model, the state space reconstructed from a signal feature is partitioned into various clusters using a Dirichlet process [18], such that each cluster follows a multivariate Gaussian distribution. A local GP expert f can be fitted for each cluster. A GP expert model [17] seeks a mapping $\vartheta = f(x) + \varepsilon$, where $x \in R^d$ is the input for the prediction model comprised of historic realizations of an extracted feature, $\vartheta \in R$ is the output (future feature), and $\varepsilon \sim N(0, \sigma_{noise}^2)$. In DP, the symmetric prior assignment probability for K clusters can be given as [18]

$$p(\boldsymbol{\pi}) = p(\pi_1, \pi_2, \dots, \pi_K|\alpha) \sim Dir\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right) = \frac{\Gamma(\alpha)}{\Gamma\left(\frac{\alpha}{K}\right)^K} \prod_k \pi_k^{\frac{\alpha}{K}-1}, \qquad \text{Eq. 6-1}$$

where $\alpha > 0$ is a concentration parameter. This is conjugate to the multinomial distribution of cluster indicators of n data points, $p(c_1, c_2, \dots, c_n|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{n_k}$ (where n_k is the number of data points in each cluster, $\sum n_k = n$). The following holds for the posterior distribution: $p(\boldsymbol{\pi}|c_1, c_2, \dots, c_n) \propto p(c_1, c_2, \dots, c_n|\boldsymbol{\pi})p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha+n)}{\prod_{k=1}^{K}\Gamma\left(\frac{\alpha}{K}+n_j\right)} \prod_{k=1}^{K} \pi_k^{\frac{\alpha}{K}+n_k-1}$.

That said, $p(\boldsymbol{\pi}|c_1, c_2, \dots, c_n) \sim Dir(\frac{\alpha}{K} + n_1, \dots, \frac{\alpha}{K} + n_K)$, and as $K \to \infty$, we can update the posterior indicator distribution using Gibbs sampling,

$$\begin{aligned} p(c_i = k|\boldsymbol{c}_{-i}, \alpha) &\propto \frac{n_{-i,k}}{n-1+\alpha} \\ p(c_i \neq c_k \forall k \neq i|c_{-i}) &\propto \frac{\alpha}{n-1+\alpha} \end{aligned} \quad \text{Eq. 6-2}$$

where $n_{-i,k}$ is the number of data points in cluster k before the assignment of data point i. The distribution for a new input within a mixture cluster $p(x_*|c = k, \mu_k, \Sigma_k) \sim N(\mu_k, \Sigma_k)$. Here, the parameters μ_k and Σ_k are the mean and the covariance for cluster k, with Gaussian distribution and an inverse Wishart distribution prior, respectively. We can obtain the weight for each cluster as:

$$w_k = p(c = k|x_*) = \frac{p(x_*|c=k)p(c=k)}{\sum_{k=1}^{K} p(x_*|c=k)p(c=k)}. \quad \text{Eq. 6-3}$$

For a realized signal feature (i.e., input) x_*, we can obtain local predictions $\overline{\vartheta}_k$ $(k = 1,2, \dots, K)$ from each local GP expert,

$$\begin{aligned} \overline{\vartheta}_k &= K(X_k, x_*)^T \left[Q(X_k, X_k) + \sigma_{noise}^2 I\right]^{-1} \zeta_k \\ cov(\vartheta_k) &= Q(x_*, x_*) - \\ Q(X_k, x_*)^T &\left[Q(X_k, X_k) + \sigma_{noise}^2 I\right]^{-1} Q(X_k, x_*) \end{aligned} \quad \text{Eq. 6-4}$$

Here, $\mathrm{X_k} = \left[\mathrm{x_1^k}, \dots, \mathrm{x_{n_k}^k}\right]^{\mathrm{T}}$ and $\zeta_\mathrm{k} = \left[\vartheta_1^\mathrm{k}, \dots, \vartheta_{\mathrm{n_k}}^\mathrm{k}\right]^{\mathrm{T}}$, are the observation samples (input and output) in cluster k, and Q is the covariance matrix defined in terms of a covariance function [17]. Then the prediction for input x_* can be expressed as a weighted sum

$$\vartheta_* = \sum_{k=1}^{K} w_k \bar{\vartheta}_k \quad \text{Eq. 6-5}$$

For multi-step predictions, after the first step, the input to the DPMG model is a Gaussian random vector, as obtained from previous-step prediction. We assume the new input $x_* \sim N(\mu_{x_*}, \Sigma_{x_*})$, where μ_{x_*} and Σ_{x_*} can be estimated from Eq. 6-4. The output distribution in each cluster is given by

$$P\left(f(x_*)\middle|\mu_{x_*}, \Sigma_{x_*}, X_k, \theta_k\right) = \int P(f(x_*)|x_*, X_k, \theta_k) P(x_*) dx_* \quad \text{Eq. 6-6}$$

The integration in Eq. 6-6 is a complicated function of x_*, a closed form expression for the output distribution was not sought, and we used a Monte Carlo approach to approximate the expression at the right side of (6) as:

$$P\left(f(x_*)\middle|\mu_{x_*}, \Sigma_{x_*}, X_k, \theta_k\right) = \frac{1}{T}\sum_{t=1}^{T} P(f(x_*^t)|x_*^t, X_k, \theta_k) \quad \text{Eq. 6-7}$$

where $T = 2000$ is the total number of random samples. The expert function is evaluated with samples from the random input distribution $N(\mu_{x_*}, \Sigma_{x_*})$, and the average is used as the predicted feature value. The performance of the DPMG model for multi-step ahead prediction of sleep apnea was compared as part of the following validation study with that from the Autoregressive Moving Average (ARMA) and Empirical Mode Decomposition (EMD) prediction techniques [19].

6.3.3 Clinical validation

The two sources of data used in this research are collected from the Apnea-ECG Database – Physionet.org and from the wireless multisensory platform developed by COMMSENS (OkState) lab. The first source of data is 20 recordings that include an ECG signal sampled at 100Hz, 16 bit resolution, synchronized with a set of minute-wide apnea annotations. The annotations of sleep apnea are made by a human expert and are based on supplementary signals including chest and abdominal respiratory effort, oronasal airflow, and oxygen saturation signals. The second source of data was collected from six healthy male subjects and two subjects with suspected sleep apnea

(age range of 25-40) using the wireless wearable multisensory suite. The subjects participating in this research were trained to use the devices and the corresponding software with professional technicians at the COMMSENS lab. However, because of the nature of the sleep testing procedure, the data collection processes were performed entirely at the subject's home. In addition to the suite, a portable sleep recording device (from Zeo) with automated algorithms to distinguish between sleep and wakefulness stages [20], was used to record the sleep stages and rate the sleep quality using an average sleep score. Both the multisensory suite and the portable sleep monitoring device wirelessly record sleep stages and provide sleep and wakefulness patterns that are quantified into four stages of sleep: wakefulness, rapid eye movement (REM) sleep, light sleep (combined Stages 1 and 2 of sleep), and deep sleep (combined Stages 3 and 4 of sleep) [21]. The portable sleep recording device with a subject wearing the multisensory suite is shown in Figure 6-4. Software with an appropriate graphical user interface was provided to assist subjects with data collection.

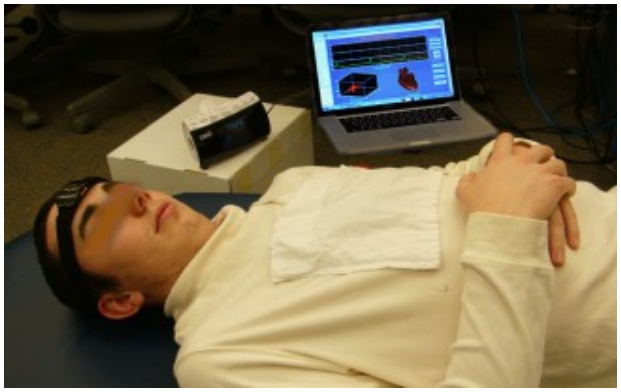


Figure 6-4 A multisensory suite with portable sleep monitoring device

The procedure for clinical validation in this study consists of collecting signals from the multisensory suite and the portable sleep device for two consecutive nights from each subject. For the first experimental epoch (sleep through one night), each subject was requested to use only the

sleep device (from Zeo) during sleep. The sleep pattern and quality score from the sleep device were collected. The purpose of this experiment is to collect the sleep quality score of subjects without wearing the multisensory suite, for validation purposes. For the second experimental epoch (i.e., the second night), the subject(s) donned the multisensory suite and used the sleep device. The signals from both devices were collected in real-time and saved in a secure location in a computer accessible via a relational database management system. The purpose of this experiment was to validate the effect of wearing the wireless multisensory suite on sleep quality, as well as to glean circumstantial estimates of the accuracy (sensitivity and specificity) of the predictions compared against the sleep events recorded by the sleep device and the signals picked up by the sound sensor.

6.4 Implementation details and results

6.4.1 Feature extraction

First, a band-pass filter with a pass band in the range of 0.06-40Hz was employed to remove the noise, artifacts, and base-line wandering and retain the critical features for the R peak extraction from the VCG signals. After de-noising, the R peaks of the ECG signal are detected by using wavelet transformation [22]. The heart rate time series known as RR intervals is calculated as the time difference between consecutive R peaks. Abnormal heart rates characterized by at least 80% increase over the previous beat's are eliminated. The power spectral density (PSD) of the RR intervals in a low frequency band (0.04 to 0.12 Hz) is used to capture the heart rate variability in OSA patients. The PSD time series is formulated such that each point is the average power spectral density of one minute of the RR interval time series. The normalized PSD (NPSD) feature is considered to account for inter-subject variability.

Recurrence quantification analysis (RQA) is employed to capture the nonlinear and nonstationary characteristics of the RR interval signals. The time delay $\zeta = 5$, which is determined based on the mutual information test [23], and the embedding dimension $d = 7$, based on the false nearest neighbors test [24], were used to reconstruct the phase space. The threshold of the recurrent plot is identified as 10% of the maximum phase space diameters [25]. The RQA features are extracted based on a sliding window concept with a window size of 600 data points and a sliding step of 60 data points corresponding to a 10 min length and a 1 min step of the RR interval time series, respectively. The 10 min length for the sliding window was selected to accommodate the longest likely sleep apnea episode a subject may experience. The sliding step of 60 sec is sufficient to characterize the cyclic variance of the heart rate which ranges from 20 to 60 sec. The recurrence features of each sliding window extracted from the recurrence plot of the 10 min RR interval in the phase space qualify for the recurrent characteristics of heart rate variability in OSA patients. The features extracted from the recurrence plot are the recurrence rate (RR), determinism (DET), average length of the diagonal lines (DIA), length of the longest diagonal line (LMAX), entropy (ENT), laminarity (LAM), trapping time (TT), length of longest vertical line (LVM), recurrence time of 1st type (RT1), recurrence time of 2nd type (RT2), recurrence period entropy density (RENT), and transitivity (TRAN).

6.4.2 Classification model

We employed a nonlinear support vector machine (SVM) classification model to discern sleep apnea events based on the extracted PSD and RQA features. An SVM classifier separates a set of binary labeled training data with a maximal margin hyper-plane, i.e., it is oriented as far away as possible from the closest members of both classes (known as the support vectors) [26]. If no linear separation exists, the original input space of features $\boldsymbol{x_i}$ is transformed to another

isomorphic space where the training set becomes linearly separable or mostly linearly separable. We need not compute the transform $\varphi(.)$ explicitly; instead we only need to estimate the inner product of the mapped patterns, $\boldsymbol{k}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \langle \varphi(\boldsymbol{x_i}), \varphi(\boldsymbol{x_j}) \rangle$ where $\langle \cdot \rangle$ denotes the inner product. The inner product is expressed as a linear combination of specified kernel functions. Based on the kernel function used, SVMs are categorized into linear, (Gaussian) radial basis function (RBF), polynomial, and multilayer perceptron classifiers. In the present work, we selected the Gaussian RBF kernel based on its superior performance over other classifiers in K-fold cross validation studies.

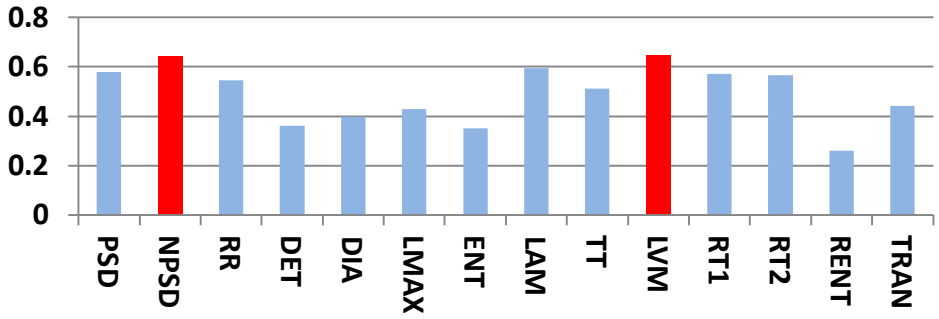


Figure 6-5 KS statistic indicates the maximal feature distribution differences between sleep apnea and non-apnea groups.

To reduce the high dimensionality of the input space (14 features), the features that most effectively classify the input space into sleep apnea and non-apnea groups were selected based on the Kolmogorov−Smirnov (KS) statistic. Figure 6-5 shows the KS statistic values of 14 features. The two significant features with the highest KS statistic— NPSD and LVM —were selected as the inputs of the classifier. Figure 6-6 (a) shows the distribution of the sleep apnea and nonapnea events in the 2D feature space of NPSD and LVM and Figure 6-6 (b) the classification boundary using Gaussian RBF. It may be noted that the nonapnea feature values are clustered in the green dot circle with low NPSD and LVM values. Table 6-1 summarizes the sensitivity and specificity

of the classification with different percentages of training data. It is noted that the average sensitivity and specificity for all classification cases are above 75% with standard deviation < 6% even with 10% of the data for training. With 90% of data used for training, the sensitivity and specificity increased remarkably to 92.56% and 86.92%, respectively. The high classification accuracies possible from the SVM model allows the use of the feature values predicted from the DPMG model to forecast the onset of an impending apneic event.

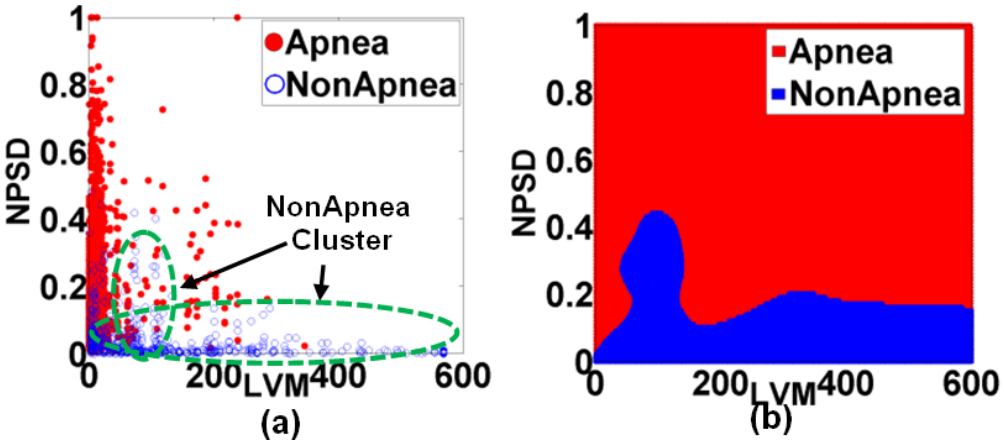


Figure 6-6 a) Distribution of apnea and nonapnea events in 2D feature space (NPSD and LVM). b) The classification boundary of the selected Gaussian RBF kernel used as part of the SVM classifier.

Table 6-1 Comparison of the accuracy (sensitivity and specificity) of Support Vector Machine classification at different training levels

Acc. (%) / Training	Apnea		NonApnea	
	Mean	**Std.**	**Mean**	**Std.**
10%	80.82	5.63	75.52	4.45
50%	86.75	3.24	82.45	3.25
90%	92.56	2.88	86.92	2.24

6.4.3 Prediction results

Among the prediction methods tested, DPMG yields the highest R^2 and classification accuracy for different prediction horizons as summarized in Table 6-2. Here we quantified the performance of the feature predictions in terms of the R^2 statistic, and the performance of overall apnea event

forecasts in terms of classification accuracy. It is noted that the DPMG model performs better than the classical ARMA and EMD models both in prediction and classification. Furthermore, when the prediction horizon increases, the accuracy of the DPMG model does not drop significantly. Figure 6-7 shows the training and prediction data of the LVM, NPSD features and the sleep apnea status with the prediction point started at the 341st min. It is observed that the DPMG model with different prediction horizons can capture the trend and the amplitude of the observation features. Thus it yields reasonably high prediction accuracies of apnea conditions (i.e., 83% for 1 step-ahead prediction and 77% for 3 step-ahead predictions).

Table 6-2 Comparison of the accuracies for 1 min and 3 min look-ahead predictions of OSA episodes with different models

Method	R^2 (first/last Step)	Classification accuracy (first/last step)
ARMA	0.37/0.1	0.4/0.03
EMD	0.45/0	0.67/0.53
DPMG	0.92/0.51	0.83/0.77

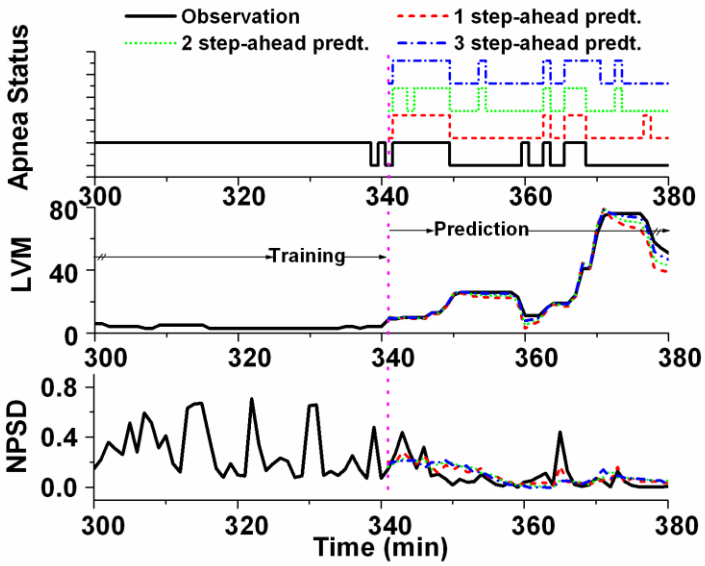


Figure 6-7 Observation from 300th to 380th min and multiple step-ahead predictions from 341th to 380th min of sleep apnea status, LVM, and NPSD features from patient a05.

Table 6-3 summarizes the average percentage of the total sleep time spent in each of the four stages of the sleep, namely, awake, rapid eye movement (REM), light sleep and deep sleep. The average was computed using the recordings of the portable sleep device taken from 8 subjects over two nights with and without donning the multisensory suite. It may be noted from the table that no significant changes exist in any of the four sleep stages between the 'No Suite' and 'With Suite' cases. This indicates that the multisensory suite may not adversely affect the quality of the sleep from a "comfort of wear" standpoint. All the eight subjects who have participated in this study have also affirmed that they found it comfortable to wear the suite to sleep.

Table 6-3 Comparison of the average percentage of time durations in four stages of sleep with and without adorning the wearable multisensory suite

Duration (%)	Wake	REM	Light	Deep
No suite	0.49	34.75	52.36	12.39
With Suite	0.74	38.29	49.80	11.17

Figure 6-8 shows a representative 350 min long signal segments collected from a subject who donned the multisensory suite to sleep for 434 minutes. This subject suspects that he suffers from sleep apnea and showed several signs of sleep apnea including loud snoring and disturbed sleep. The signals shown in the figure include synchronously gathered sound (black), sleep stage (blue) signals along with the online one-minute-ahead sleep apnea predictions (red). The start of the signal strip coincides with the time when the subject begins to sleep, as indicated by the transition from wake to light sleep stage. It may be noted that no apneic episode was predicted during the 25 min long deep sleep, or during the stable light sleep stage in the 200-300 min range. More pertinently, the first apnea event predicted at the 60 min mark precedes the transition from a deep to light sleep. Also the 2^{nd}, 3^{rd}, 4^{th}, 6^{th}, 7^{th}, and 9^{th} apnea events are predicted near the transitions from a deep sleep to a light or REM sleep. Two apparent false positive apnea predictions (events 8^{th} and 10^{th}) occur at 225^{th} min and 336^{th} min marks in REM sleep stage.

Offline training epochs to estimate the parameters of the classification and prediction models using the longitudinal data collected from a subject might further improve the sensitivity and specificity of the predictions.

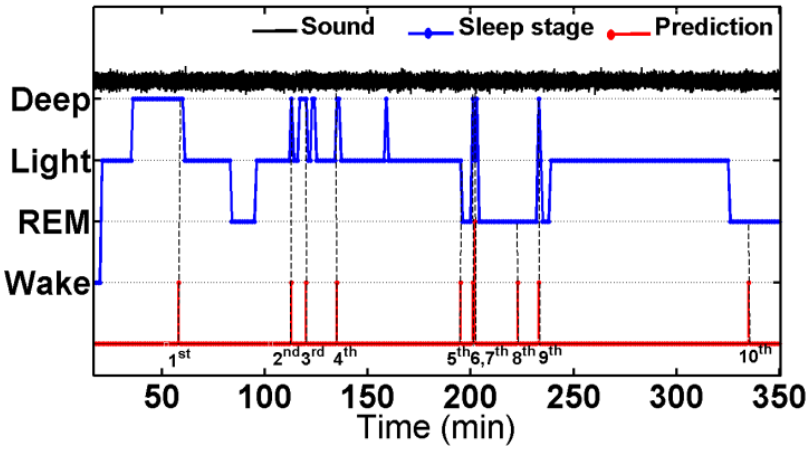


Figure 6-8 Real-time sound signal, sleep stage pattern, and one-minute ahead prediction of sleep apnea in subject ID008 from the starting of sleep to 350th min

6.5 Conclusions

We have developed an approach to provide 1-3 min ahead early warning of an impending sleep apnea episode based on using a wearable wireless multisensory suite and a novel nonlinear nonstationary process prediction method. The wearable wireless multisensory system can serve as a viable platform to continuously and noninvasively acquire physiological signals to track cardiorespiratory dynamics, and quantitatively assess apneic conditions for prediction of OSA episodes. Testing of the wearable sensory suite among eight mostly healthy subjects suggests that our sensory suite does not adversely impede the comfort and quality of sleep. The prediction approach was extensively tested using 20 recordings from the Physionet database and 10 recordings from 8 subjects wearing the multisensory unit. These tests indicate that the classification and prediction accuracies (R^2) of 70-90% are possible from the present approach. It was also evident that the longest vertical length (LVM) of the recurrence plot and normalized power spectral density (NPSD) are the most sensitive features for OSA episode prediction with

offline OSA classification accuracy of up to 88%. More pertinently, DPMG was shown to provide OSA prediction accuracy of 83% 1 step-ahead and 77% for 3 step-ahead. This amounts to some 20-40% improvement in prediction accuracy compared to other methods tested. Such early prediction can spur the development of adaptive flow control systems for CPAP devices and to the advent of devices to induce minor adjustments to body positions to mitigate OSA.

The ongoing investigations are focused on testing the approach on additional subjects to extract consistent benchmarks of the performance of the multisensory suite and the prediction method. Also, it may be noted that the predictions reported in the present work are based on the SVM classification model derived from multiple subjects. We are currently investigating the customization of the classification and predictions to subjects, which we believe can further improve the performance of our approach. Additionally, we are pursuing methods to mitigate OSA episodes that take advantage of the early prediction capability.

REFERENCES

[1] T. Young, *et al.*, "The Occurrence of Sleep-Disordered Breathing among Middle-Aged Adults," *New England Journal of Medicine,* vol. 328, 1230-1235, 1993.

[2] P. Lavie, *et al.*, "All-Cause Mortality in Males with Sleep Apnoea Syndrome: Declining Mortality Rates with Age," *European Respiratory Journal,* vol. 25, 514-520, 2005.

[3] L. S. Doherty, *et al.*, "Long-Term Effects of Nasal Continuous Positive Airway Pressure Therapy on Cardiovascular Outcomes in Sleep Apnea Syndrome," *Chest,* vol. 127, 2076-2084, 2005.

[4] J. M. Marin, *et al.*, "Long-Term Cardiovascular Outcomes in Men with Obstructive Sleep Apnoea-Hypopnoea with or without Treatment with Continuous Positive Airway Pressure: An Observational Study," *The Lancet,* vol. 365, 1046-1053, 2005.

[5] T. E. Weaver and R. R. Grunstein, "Adherence to Continuous Positive Airway Pressure Therapy," *Proceedings of the American Thoracic Society,* vol. 5, 173-178, 2008.

[6] L. Hood, *et al.*, "Systems Biology and New Technologies Enable Predictive and Preventative Medicine.," *Science,* vol. 306, 640-643, Oct 22 2004.

[7] Q. Tian, *et al.*, "Systems Cancer Medicine: Towards Realization of Predictive, Preventive, Personalized and Participatory (P4) Medicine," *Journal of Internal Medicine,* vol. 271, 111-121, 2012.

[8] R. Snyderman and Z. Yoediono, "Prospective Care: A Personalized, Preventative Approach to Medicine.," *Pharmacogenomics,* vol. 7, 5-9, Feb 2006.

[9] J. E. Stahmann, *et al.*, "Prediction of Disordered Breathing," United States Patent US 7938782 B2, 2011.

[10] T. Penzel, *et al.*, "Systematic Comparison of Different Algorithms for Apnoea Detection Based on Electrocardiogram Recordings," *Medical and Biological Engineering and Computing,* vol. 40, 402-407, 2002.

[11] M. O. Mendez, *et al.*, "Automatic Screening of Obstructive Sleep Apnea from the Ecg Based on Empirical Mode Decomposition and Wavelet Analysis," *Physiological Measurement,* vol. 31, 273-289, 2010.

[12] O. Fontenla-Romero, *et al.*, "A New Method for Sleep Apnea Classification Using Wavelets and Feedforward Neural Networks," *Artificial Intelligence Medicines,* vol. 34, 65-76, May 2005.

[13] P. Dagum and A. Galper, "Time Series Prediction Using Belief Network Models," *International Journal of Human Computer Studies,* vol. 42, 617-632, 1995.

[14] J. Bock and D. A. Gough, "Toward Prediction of Physiological State Signals in Sleep Apnea," *IEEE Transactions on Biomedical Engineering,* vol. 45, 1332-1341, 1998.

[15] E. D. Weitzman, *et al.*, "Quantitative Analysis of Sleep and Sleep Apnea before and after Tracheostomy in Patients with the Hypersomnia-Sleep Apnea Syndrome," *Sleep,* vol. 3, 407-423, 1980.

[16] S. T. S. Bukkapatnam, *et al.*, "Wireless Multi-Sensor Platform for Continuous Real-Time Monitoring of Electrocardiogram Signals," 2010.

[17] S. T. S. Bukkapatnam and C. Cheng, "Forecasting the Evolution of Nonlinear and Nonstationary Systems Using Recurrence-Based Local Gaussian Process Models," *Physical Review E,* vol. 82, 056206, Nov 2010.

[18] D. M. Blei and M. I. Jordan, "Variational Inference for Dirichlet Process Mixtures," *Bayesian Analysis,* vol. 1, 121-144, 2005.

[19] A. Sa-ngasoongsong and S. T. S. Bukkapatnam, "Willingness-to-Pay Prediction Based on Empirical Mode Decomposition," presented at the Industrial Engineering Research Conference, Reno, Nevada, 2011.

[20] J. R. Shambroom, *et al.*, "Validation of an Automated Wireless System to Monitor Sleep in Healthy Adults," *Sleep Research,* vol. 21, 221-230, 2012.

[21] E. A. Wolpert, "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects," *Archives of General Psychiatry,* vol. 20, 246-247, 1969.

[22] V. X. Afonso, *et al.*, "Ecg Beat Detection Using Filter Banks," *IEEE Transactions on Biomedical Engineering,* vol. 46, 192-202, 1999.

[23] A. M. Fraser and H. L. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information," *Physical Review A,* vol. 33, 1134-1140, 1986.

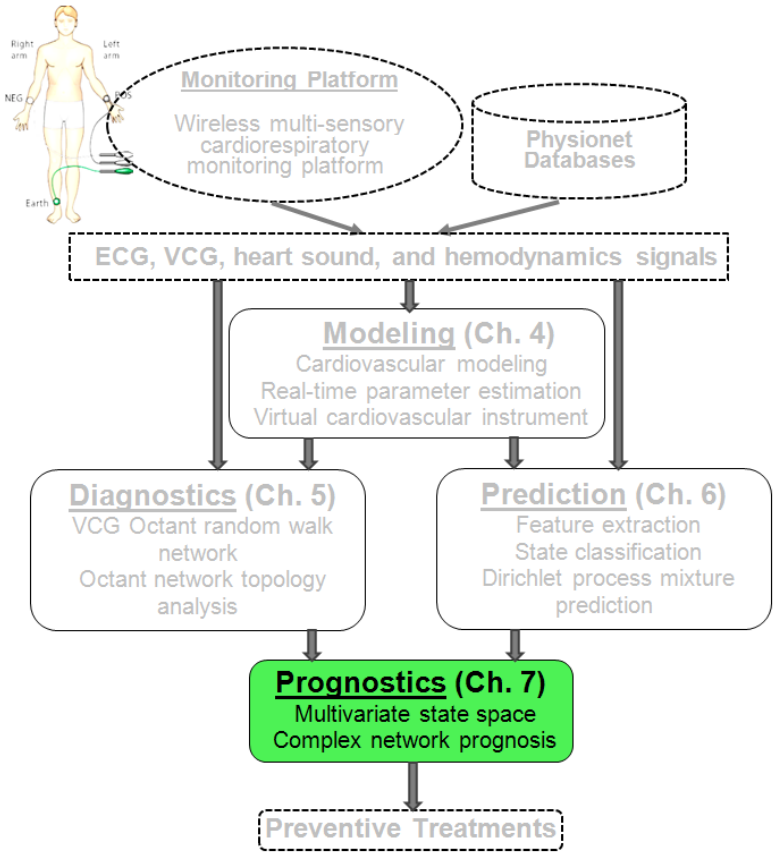
[24] M. B. Kennel, *et al.*, "Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction," *Physical review A,* vol. 45, 3403-3411, 1992.

[25] H. Kantz*, et al.*, *Nonlinear Time Series Analysis* vol. 2000: Cambridge University Press Cambridge, 1997.

[26] A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," *Statistics and Computing,* vol. 14*,* 199-222, 2004.

CHAPTER 7

NONPARAMETRIC MODELING APPROACH FOR PERSONALIZED PROGNOSIS OF CARDIORESPIRATORY DISORDERS



This chapter describes an approach to deriving personalized estimates of remaining useful life (RUL) distributions (e.g., time till an abnormal condition) based on the diagnostic method and Dirichlet process mixture prediction model presented in chapters 5 and 6, respectively. The prognostic approach is based on deriving a state space network representation of the pathways of

transition from various healthy states to abnormal states. The prognostic model is validated using data from the ECG Apnea Database (Physionet.org). The average prediction accuracy (R^2) is reported as 0.75%, with 87% of observations within the 95% confidence interval. Estimated risk indicators at 1 to 3 min till apnea onset are reported as $85.8 \pm 5.7\%$, $80.2 \pm 10.1\%$, and $75.5 \pm 10.8\%$, respectively.

7.1 Introduction

Chronic disorders such as obesity are considered an epidemic afflicting nearly 133 million Americans and utilizing 75% of the nation's healthcare spending [1]. The transition of obese patients into other chronic conditions, such as obstructive sleep apnea (OSA) [2-4] and life-threatening cardiovascular diseases (CVDs) is particularly alarming [5]. Approximately 70% of OSA patients are obese [6], and 4% with a CVD [7], such as atrial fibrillation and hypertension. Therefore, the continual monitoring of health status and assessment of health risks for each individual are necessary to help mitigate transition into more serious chronic or acute disorders, and to potentially reverse some chronic disorders.

Growing healthcare costs and recent technological advancements such as those in genomics and wireless sensors are fueling the desire to transform healthcare practice from reactive "damage control" to a more proactive and personalized wellness process [8-10]. The emphasis on disease prevention, early detection, and preventive treatments will revolutionize the way clinicians evaluate their patients. Current medical prognostic practice relies primarily on comparison to "population norms" extracted from certain genetic and environmental factors to estimate a patient's predisposition to a disease [11]. But it is not uncommon for clinicians to encounter individuals with high predisposition to a disease who remain healthy. Population norms do not provide personalized, dynamic quantification, and the ability to forecast the risks in an individual

who is undergoing a transition to a serious chronic or acute condition, especially over the short-term.

Recent studies suggest that deterioration in a patient's health status is often preceded by remarkable changes in heart rate variability dynamics and/or coupling dynamics among the cardiovascular processes [12-14]. These changes are evident to hospital staff hours before the acute onset. The pathways of the transitions between various chronic and acute health statuses can be monitored by tracking the short term transitions and long term degradation processes underlying the physiological signal dynamics. Such transitions are difficult to discern using traditional clinical features (measures), such as heart rate and respiration profiles. However, nonlinear dynamic quantifiers rooted in nonlinear dynamic systems theory (e.g., local topology analysis and recurrence quantification analysis features) could be used not only to monitor, but also to predict transitions well before they become evident. These prediction methods would allow the forecast of diseases from months to years before observable symptoms of disease occur.

This research introduces a method to visually represent the transitional pathways of abnormal states in the dynamic state space of the underlying system and a new method using the local topology of cardiovascular dynamics to track the degradation transitions and derive the remaining useful life (RUL) distribution. The new systems dynamic representation and evolution-based prognostic method facilitate the forecasting of acute or chronic health transitions. Furthermore, accurate risk indicators derived from the distribution of RUL support the development of preventive treatments and intervention therapies for proactive and personalized wellness processes. The organization of the remainder of the chapter is as follows: section 7-2 reviews the literature on the current prognostic approaches; section 7-3 describes the overall methodology of the proposed prognostic method; implementation of the methodology and a case

study in estimating the distribution of remaining time to sleep apnea onset are presented in section 7-4; and the conclusions are provided in section 7-5.

7.2 Background and literature review

Prognostic models are intended to provide reliable predictions about the RUL of components or systems undergoing degradation processes. Prognostic models can be categorized broadly into model-based approaches and data-driven approaches. Model-based methods use an explicit mathematical model of the degradation process to predict the future evolutions of the degradation state and, thus, the RUL of the system. Data-driven methods are used when an explicit model of the degradation process is not available, but a statistical model that can capture the historical trend of the data is derivable. Biological and mechanical systems are the two major target areas for most of the research in prognostics. In this research, we have summarized the published works on prognostic models using model-based and data-driven approaches in mechanical and biological (with a focus on cardiovascular) systems.

Model-based and data-driven prognostic approaches have been receiving increasing attention for maintenance practices in mechanical systems. Model-based methods characterize the dynamic degradation state by explicitly using mathematical models of the degradation process. Pulkkinen *et al.,* [15] and Ray *et al.,* [16] proposed non-linear stochastic damage models to characterize fatigue crack dynamics. Kim and Kolarik [17] developed a real-time conditional reliability prediction for an individual component using model performance data, and formulated real-time conditional reliability predictions based on regression analysis. Lu *et al.,* [18] modeled the degradation signal of an individual product by exponential smoothing and calculated the predictive reliability based on the S-normal distribution of the performance variable. Chinnam *et al.,* [19] modeled the degradation process using a general polynomial regression with the

reliability function following a T- distribution. Jianhui *et al.,* [20] introduced an integrated prognostic process based on data from model-based simulations under nominal and degraded conditions. Chelidze and Cusumano [21] suggested a general method for tracking the evolution of hidden damage processes using the reconstructed phase space of the reference system. Many Bayesian-based prognostic approaches [22-24] modeled stochastic damage processes using the exponential degradation function for the degradation signal to update the RUL distribution with the assumption that failure occurs when the degradation signal reaches some given failure thresholds. Data-driven methods are used when historical sensory signals (i.e., vibration, cracking data) have been collected but an explicit model of the degradation process is not available. Data-driven models update the RUL distribution by building statistical models that fit the historical data. On this basis, autoregressive moving averages [25], artificial neural networks [26, 27], and relevance vector machines [25, 28] have been used to assess the distribution of the remaining time. Recently, ensemble approaches based on the aggregation of multiple model outcomes have been introduced to combine the advantages of superior robustness and accuracy with respect to single models [29, 30].

Providing medical prognostics model for biological systems plays an important role in directing disease diagnosis and treatment pathways. In this section we focus on medical prognostic approaches for cardiovascular disorders. Not many studies have been reported on model-based approaches to prognosis due to the limitations in modeling anatomical and physiological characteristics of the cardiovascular system. A typical model-based prognostic tool is the KARDIO system [31]. This is a rule-based system built from anatomical and physiological knowledge that is able to diagnose and predict the conditions of the heart from the trigger of cardiac arrhythmias in the patient. For a data-driven prognostic model, different statistical models, including Cox's proportional-hazards model [32-35], the Kaplan-Meier model [36-38],

and the Mantel log rank test [39-41] have been used to predict expected RUL. The most powerful predictors of CVD risks in the statistical models are age, gender, body mass index (BMI), blood pressure (BP), total cholesterol, LDL cholesterol, HDL cholesterol, smoking habits, diabetes, hypertension, and albuminuria [42-45]. More personalized risk assessment approaches pursued include the tracking of various types of cholesterol [46-48]; evolving BP, BMI, body shape measurements [49-51]; biomarkers such as C-Reactive Protein, and N-terminal pro-brain natriuretic peptide [52-55]; ECG analysis [56-59]; electron beam CT scans [60-62]; cardiac magnetic resonance [63-65]; and genetic variation [66, 67].

Current prognostic models are derived from limited data gathered from small population sizes; hence these models need some form of baseline adjustments before they can be generalized to different subjects. These approaches are rooted in comparisons to "population norms" and are not able to capture the physiological dynamics of individuals. Furthermore, few efforts (if any) have been reported towards using continuous tracking nonlinear transitions. Current medical practices capture data only during a short period (i.e., ambulatory signals, clinic measurements); such short-term data is not able to capture long-term degradation, which leads to less accurate risk estimation [68]. Methods based on tracking the nonlinear and nonstationary evolution of the dynamic underlying monitoring physiological signals need to be investigated to facilitate individualized treatments and timely prevention therapies.

7.3 Research approach

The contributions of the present research are (a) a representation of a transition state space network that is able to optimally separate the state vectors in regions with different dynamic characteristics and visually track the dynamic transitions of abnormal states in the state space (b) a prognostic method for estimating the distribution of the time to failure by considering

the stochastic evolution of the normal state vectors to the abnormal state in the state space of the underlying measured signal dynamics. As summarized in Figure. 7-1, we used data from the Apnea-ECG database (Physionet.org) for training and testing the classification and prognostic model.

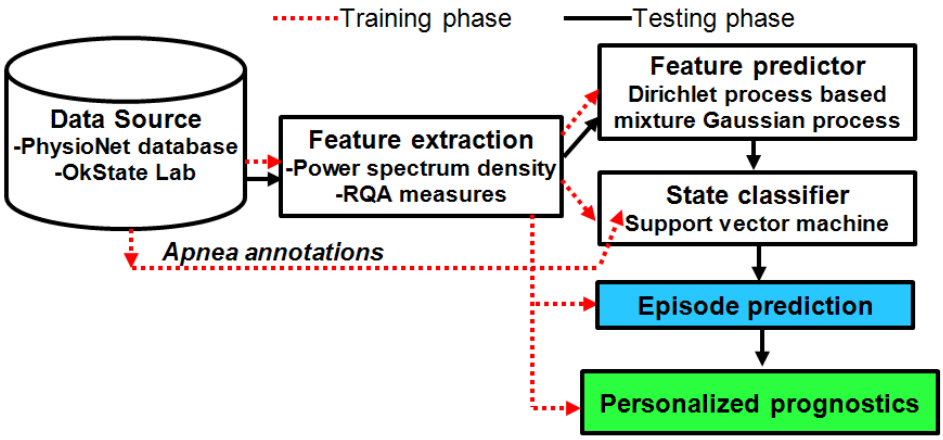


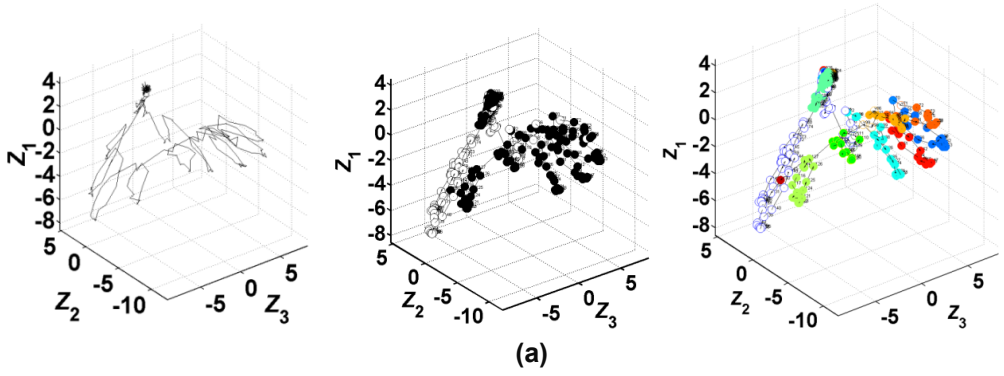
Figure 7-1 Overall approach for prognostic model

The prognostic model provided the RUL distribution and a risk indicator for the abnormality onset using the topological transitions of state space vectors reconstructed from the condition monitoring features $\boldsymbol{\vartheta}$. This work was inspired by the concept of symbolic dynamics, which allows characterizing the dynamic system based on partitioning its phase space into K mutual disjoint Dirichlet clusters. The state classification of the phase space and the evolutions of state vectors were utilized to represent a complex network with the nodes classified as state vectors and the edges as the plausible state evolutions. A Laplacian-projection method was employed to maximize cluster separation and highlight the local transitions of the state space vectors. The stochastic evolution of $\boldsymbol{\vartheta}(t)$ among clusters visually presented in the Laplacian-projection system was used to derive the distribution of time to an onset of a failure. Distribution $f_T(t|\boldsymbol{x}_*)$ of the random variable time to failure T(or RUL) is derived from the probability that in a phase space network, there exists a path from a normal state $\boldsymbol{x}_*$ to an abnormal state with specific length t.

The remainder of this section describes two main contributions of this work: a state space network representation and a prognostics approach.

7.3.1 Color coded state space representation

State space reconstruction for multivariate time series has been used for the state space network representation. The embedded state space is represented as a directed graph $G(V, E)$ (see Figure 7-2 (a)) with the node set $V = (v_t, t = 1,2, \dots n)$ as state vectors and edge set E as the plausible evolutions of the state vectors in the state space. The distances between the nodes are Euclidian distances in an M-dimensional embedding space. Figure 7-2 shows a state space network in 3-D with the nodes classified into normal and abnormal states (solid and empty circles) and grouped in different clusters with distinct colors using the Support Vector Machine (SVM) classification and Dirichlet process based mixture Gaussian process (DPMG) models [69] mentioned in chapter 6. Since the proposed prognostic method focuses on tracking the dynamic transitions of the state vectors in the network, visualization of the state transitions is very important. However, due to the high embedding dimension (M>8) of the state space and the overcrowding of adjacent state vectors in the same cluster, visualization of the state vector evolutions in the state space becomes obscured. Towards addressing these limitations, we have investigated the Laplacian-eigen projection method for the visualization of state space vector transitions in a 2-D coordinate system.



(a)

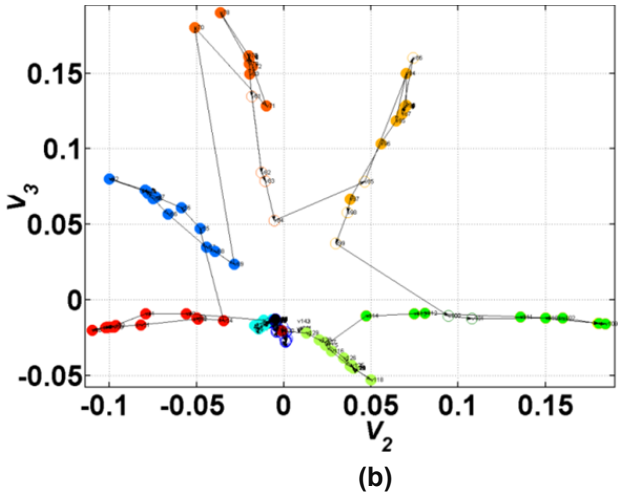


Figure 7-2 Graph representation of the reconstructed state space

The Laplacian-eigen projection method provides two distinct advantages that legitimate the layout limitations of the state space network. First, it provides mathematical formulations of the Non-deterministic Polynomial-time hard (NP-hard) problem leading to a methodical solution for the network layout. The second advantage is the fast computational speed in representing a big network with thousands of nodes in real life such as the World Wide Web, social networks, and transportation hub systems. The Laplacian matrix is defined as $L = D - \mathrm{A}$, where D is the degree matrix and A is the adjacent matrix associated with the graph $G(V, E)$. Projection of state vectors in the 2nd and 3rd smallest Laplacian-eigen vectors is subjected to a force-directed strategy [70] from which adjacent space vectors are projected more closely without overcrowding the nodes. Figure 7-2 (b) shows the projection of the state space in the Laplacian-eigen vector coordinate system. It is noted that the state vectors are separately grouped in different clusters and the transitions among adjacent state vectors are perceptible in each cluster. This Laplacian-eigen projection representation is used for further analysis of the prognostic model.

7.3.2 Prognostic model

The stochastic transitions of the state vector in the state space network have been used to derive the distribution of RUL. Figure 7-3 depicts the structure of a subgraph containing a transition path (red line) from normal state v_{t_0} in cluster C_m to abnormal state v_{t_5} in cluster C_{m+1}. Each Dirichlet cluster C_i is represented by a distinct color. Solid nodes and empty nodes represent state vectors with abnormal and normal conditions, respectively. An onset of failure is detected when a transition from an empty node to a solid node occurs. The posterior distribution of the indicator is given as:

$$p(\boldsymbol{\pi}|c_1, c_2, \dots, c_n) \sim Dir\left(\frac{\alpha}{K} + n_1, \dots, \frac{\alpha}{K} + n_K\right) \quad \text{Eq. 7-1}$$

where $c_1, c_2, \dots, c_N$ are the historical observations, $\alpha > 0$ is the concentration parameter, n_i is the number of data points in cluster C_i, and K is the number of clusters in the state space. The probability density function of a new input, $\boldsymbol{x}_*$, belonging to cluster C_m, is given as $p(\boldsymbol{x}_*|C_m, \mu_m, \Sigma_m) \sim N(\mu_m, \Sigma_m)$ where μ_m,and Σ_m are the mean and covariance of cluster C_m. The probability that $\boldsymbol{x}_*$ is within cluster C_m is given as $w_m = p(C_m|\boldsymbol{x}_*) = \frac{p(\boldsymbol{x}_*|C_m)p(C_m)}{\sum_{i=1}^{K} p(\boldsymbol{x}_*|C_m)p(C_m)}$ and the distances from x_* to state vector h are $d_h^m = \|\boldsymbol{x}_* - \boldsymbol{x}_h^m\|$.

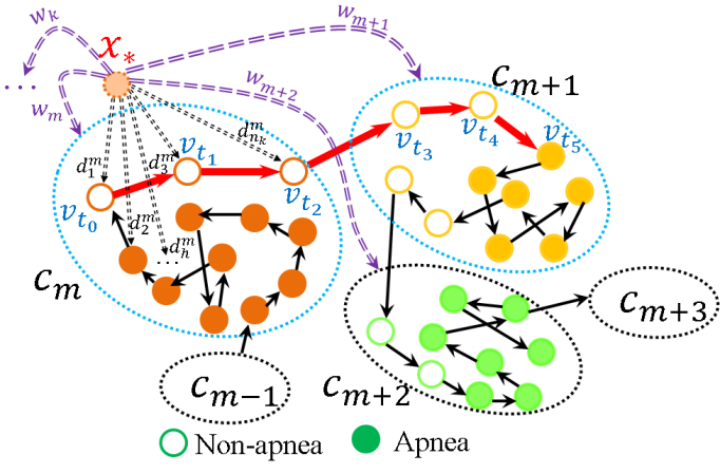


Figure 7-3 Structure of a subgraph with a transition path from a normal to an abnormal state

The RUL distribution at $\boldsymbol{x}_*$ is updated based on the distribution of the transition time for $\boldsymbol{x}_*$ to first reach an abnormal state. An abnormal state can be reached from a normal state $\boldsymbol{x}_*$ starting in cluster C_m $(m = 1, \dots, k)$ with cluster sequence length from 1 to n_{max}. Figure 7-4 lists all of the possible cluster paths that an abnormal state can be reached from a normal state.

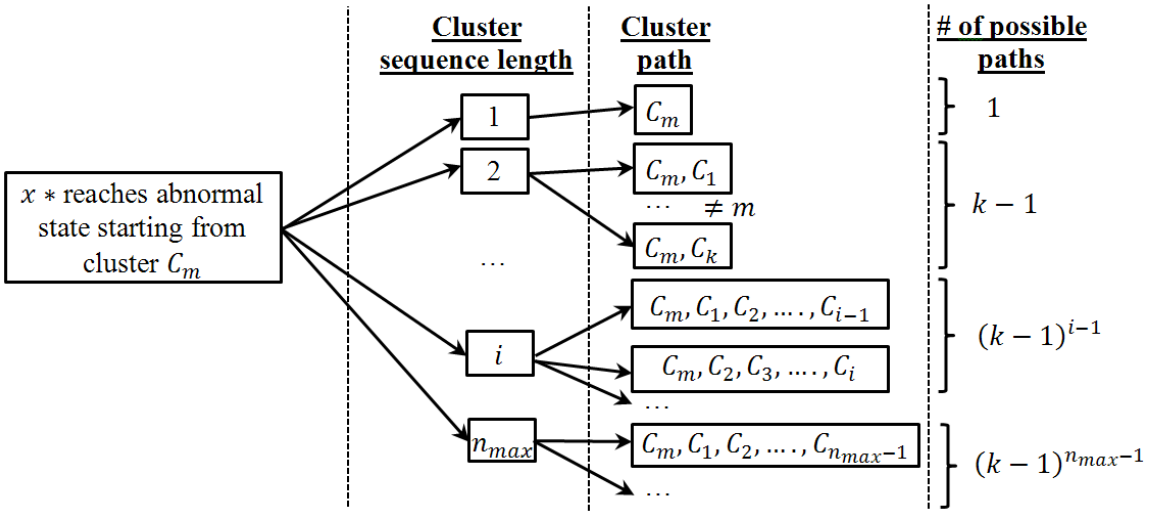


Figure 7-4 All possible paths to abnormal state from a normal state in cluster C_m

The conditional probability distribution of transition time T for a new normal condition input $\boldsymbol{x}_*$ within cluster C_m to reach the first failure onset is given as:

$$f_T^{C_m}(t|\boldsymbol{x}_*, m) = P(T = t|\boldsymbol{x}_*, v_1 = m) = \sum_{i=1}^{i=n_{max}} \phi_T^{m,i}(t) \qquad \text{Eq. 7-2}$$

where i is the cluster sequence length and v_i is the cluster indicator. The probability density function of the time to the next anomalous onset given the normal input $\boldsymbol{x}_*$ in cluster C_m and the cluster sequence i is computed recursively as:

$$\phi_T^{m,1}(t) = \sum_{\substack{k=1 \\ m \neq k}}^{K} P(v_2 = k|v_1 = m) \times P(t|v_1.v_2 = m.k) \qquad \text{Eq. 7-3}$$

$$\phi_T^{m,i}(t) = \sum_{k_1=1, k_1 \neq m}^{K} \sum_{k_2=1, k_2 \neq k_1}^{K} \cdots \sum_{k_{i-1}=1, k_{i-1} \neq k_{i-2}}^{K} P(v_2.v_3 \dots v_i = k_1.k_2 \dots .k_{i-1}|v_1 = m) \times P(t|v_1.v_2 \dots v_i = m.k_1 \dots .k_{i-1})$$

The nonparametric density of time to failure is updated as [71]:

$$f_T(t|\boldsymbol{x}_*) = \int_{\theta=C_1}^{C_k} f_T^{\theta}(t|\boldsymbol{x}_*, \theta) G(\theta)\, d\theta, \quad \text{Eq. 7-4}$$

where $f_T^{\theta}(t|\boldsymbol{x}_*, \theta)$ is the density of t given input $\boldsymbol{x}_*$ in the cluster indexed by θ and $G \sim DP(\alpha, H)$, α is the concentration parameter, and H is the inverse Wishart prior. The estimated density of time to failure is expressed as a weighted sum

$$f_T(t|\boldsymbol{x}_*) = \sum_{m=1}^{k} w_k f_T^{C_m}(t|\, \boldsymbol{x}_*, v_1 = m). \quad \text{Eq. 7-5}$$

The RUL distribution in Eq. 7-5 is used to derive the expected time to failure onset and $f_T(t|\boldsymbol{x}_*)$ is used as the risk indicators.

7.4 Implementation details and results

7.4.1 Feature extraction and classification model

Two groups of features have been used to differentiate sleep apnea states from normal ones. The first group of features is the power spectral density (PSD) of the RR intervals in a low frequency band (0.04 to 0.12 Hz).This feature captures the heart rate variability in OSA patients. The PSD time series is formulated such that each point is the average power spectral density of one minute of the RR interval time series. The normalized PSD (NPSD) feature is considered to account for inter-subject variability. The second feature group is the recurrence quantification analysis (RQA) quantifiers of the RR interval signals. RQA is employed to capture the nonlinear and nonstationary characteristics of the RR interval signals. The features extracted from the recurrence plot are the recurrence rate (RR), determinism (DET), average length of the diagonal lines (DIA), length of the longest diagonal line (LMAX), entropy (ENT), laminarity (LAM), trapping time (TT), length of the longest vertical line (LVM), recurrence time of 1st type (RT1), recurrence time of 2nd type (RT2), recurrence period entropy density (RENT), and transitivity (TRAN).

Table 7-1 Comparison of the accuracy (sensitivity and specificity) of Support Vector Machine classification at different training levels

Acc. (%) / Training	Apnea		NonApnea	
	Mean	**Std.**	**Mean**	**Std.**
10%	80.82	5.63	75.52	4.45
50%	86.75	3.24	82.45	3.25
90%	92.56	2.88	86.92	2.24

We employed a nonlinear support vector machine (SVM) classification model and Kolmogorov−Smirnov (KS) statistic for the feature selection. The SVM separates a given set of binary-labeled (e.g., apnea or non-apnea) training data with a hyper-plane that is maximally distant from them [72]. To reduce the high dimensionality of the input space (14 features), the features that most effectively classify the input space into sleep apnea and non-apnea groups were selected based on the Kolmogorov−Smirnov (KS) statistic. The two significant features with the highest KS statistic— NPSD and LVM —were selected as the inputs of the classifier. Table 7-1 Comparison of the accuracy (sensitivity and specificity) of Support Vector Machine classification at different training levels summarizes the sensitivity and specificity of the classification with different percentages of training data. It is noted that the average sensitivity and specificity for all classification cases are above 75% with standard deviation < 6% even with 10% of the data for training. With 90% of the data used for training, the sensitivity and specificity increased remarkably, to 92.56% and 86.92%, respectively. The high classification accuracies possible from the SVM model allow the use of the feature values predicted from the DPMG model to forecast the onset of an impending apneic event. Figure 7-5 shows the data from two features, PSD and LVM, and the corresponding apnea state with a 1-min resolution. The transitions of the state vectors in the state space reconstructed from these two features have been used to derive the distribution of time to apnea onset T.

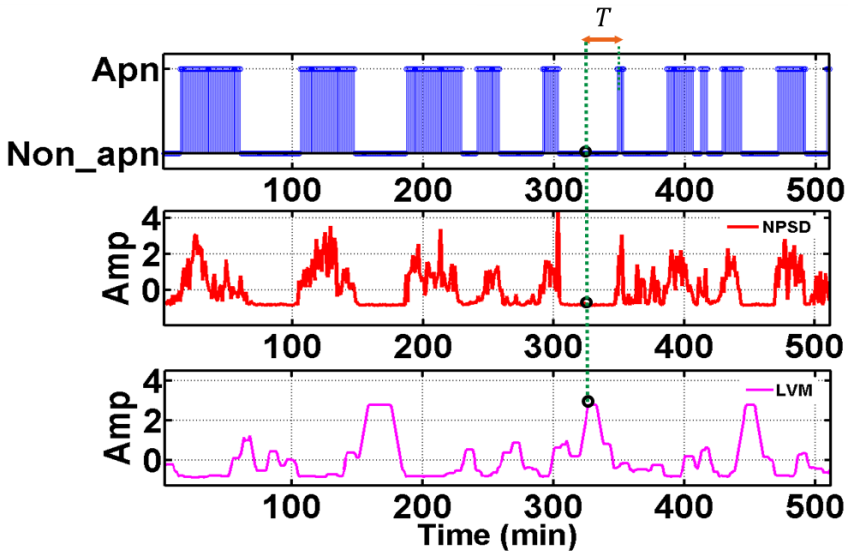


Figure 7-5 Two significant features for the state classification of sleep apnea and corresponding apnea-nonapnea annotations.

7.4.2 Multivariate time series reconstruction

Multivariate time series reconstruction of the state space was utilized to characterize the dynamic of the underlying system. Although embedding theorem [73] shows that a scalar time series is sufficient to reconstruct the system dynamics if enough information criteria are used, the situation might be different in reality. For example the measurement of the z −coordinator of the Lorenz system cannot be used to reconstruct the state space because it doesn't resolve the $x - y$ symmetry. Hence, the use of two time series provides substantial advantages in practical problems, especially in noise-driven systems.

Embedding of the state space vector of the coupled dynamic system from multivariate data is defined as:

$$\bar{Z} \equiv (x_1(t),\ x_1(t-\tau_1),\ x_1(t-2\tau_1),\ \ldots\ldots,x_1(t-(m_1-1)\tau_1),$$

$$x_2(t),\ x_2(t-\tau_2),\ldots.,x_2(t-(m_2-1)\tau_2))$$

where τ_i is the embedding time delay, m_i is the subspace dimension of x_i, and M=$\sum m_i$ is all of the embedding dimensions. The delay time, τ_i , is selected using the mutual information criteria [74]. Figure 7-6 illustrates the first three iterations to determine the multivariate embedding

dimensions using the false nearest neighbors method [75]. The method is summarized in four steps:

(1) Start with $M = 1$ (i.e. (0,1) and (1,0))

(2) Check FNN of one dimension increment (i.e., (m_1+1, m_2), (m_1, m_2+1))

(3) Terminate when the maximum norm of the FNN value drops below the threshold (i.e., $X(M_i) < \delta$)

(4) If more than one dimension vector fulfills the criterion, select the one with the smallest cumulative FNN.

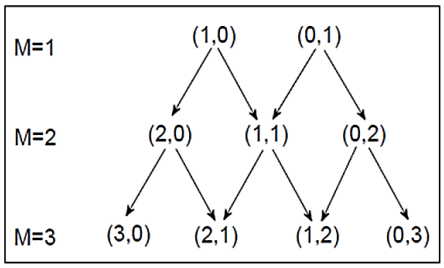


Figure 7-6 First three iterations for the determination of multivariate embedding dimensions

7.4.3 Performance of prognostics approach

A case study to estimate the distribution and expected values of time to sleep apnea onset is used to validate the prognostic method. The data set for the case study consists of multiple night recordings of 3 patients from the ECG-Apnea Database (Physionet.org). The performance of the prognostic method towards individual variances was investigated using the data from each patient. We used 90% of the data for training and 10% for testing. In the present case, the RUL distribution at $\boldsymbol{x}_*$ is updated as $f_T^{C_m}(t|\boldsymbol{x}_*, m) = P(T = t|\boldsymbol{x}_*, v_1 = m) = \sum_{i=1}^{i=n_{max}} \phi_T^{m,i}(t)$ with n_{max} =7 since all the apneic states are reached after transit through fewer than 8 clusters (as shown in Figure 7-7).

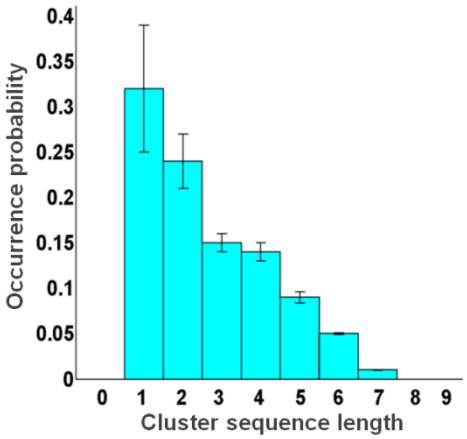


Figure 7-7 Distribution of the number of cluster transitions a non-apnea node travel to reach an apnea state

The probabilities in Eqs. (7-2 and 7-3) are estimated empirically from the state space network. The probability $P(v_2 = k|v_1 = m)$ that a failure onset is reached with the starting normal state in cluster m and the ending abnormal state in cluster k with the cluster sequence length of 2 is estimated as:

$$\hat{P}(v_2 = k|v_1 = m) = \frac{\#of\ cluster\ paths\ starting\ in\ cluster\ m\ and\ ending\ in\ cluster\ k}{\sum all\ possible\ cluster\ paths\ starting\ in\ cluster\ m\ with\ cluster\ sequence\ length\ of\ 2},$$

and the probability $P(v_2 v_3 \ldots v_i = k_1 k_2 \ldots . k_{i-1}|v_1 = m)$ that a failure onset is reached with the starting normal state in cluster m and the ending abnormal state in cluster k_{i-1} with the cluster sequence length of i is estimated as:

$$\hat{P}(v_2 . v_3 \ldots v_i = k_1 . k_2 \ldots . k_{i-1}|v_1 = m) =$$

$$\frac{\#of\ cluster\ paths\ starting\ in\ cluster\ m\ with\ cluster\ sequence\ of\ \ k_1,k_2,\ldots,and\ k_{i-1}}{\sum all\ possible\ cluster\ paths\ starting\ in\ cluster\ m\ with\ cluster\ sequence\ length\ of\ i}.$$

Finally, $\hat{P}(t|v_1 . v_2 \ldots v_i = m . k_1 \ldots . k_{i-1})$ is estimated from the probability that a first abnormal node is reached in t steps with the cluster sequence of $m, k_1 \ldots$, and k_{i-1}.

Figure 7-8 shows the observations (solid blue line) of the time to apnea onset, determined by using the sleep apnea annotations provided by experts, the expected values (red dash line) from the distribution of time to failure, and the 95% confidence interval (purple) of the prediction values. The R^2 and the percentages of observations within the 95% confidence interval of the prediction are: (0.79, 85%), (0.75, 87%), and (0.7, 90%) for the 1st, 2nd , and 3rd patient, respectively. The $R^2 \geq 0.7$ in these three cases suggests that the prognostic model can be used to predict the time to sleep apnea onset. Although the reported R^2 value of patient 3 is low, 90% of the observation values lie within the 95% confidence interval of the prediction.

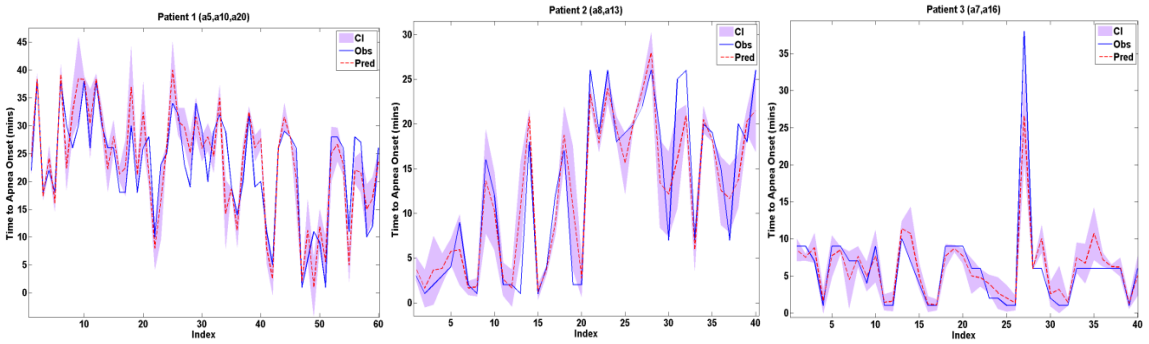


Figure 7-8 Run plot showing the predictions of the expected time to sleep apnea onset in three patients

Estimated time to apnea onset and the prediction errors are shows in: Figure 7-9 (a) 1-5 min-ahead prediction with a prediction horizon increment of 1 min, and Figure 7-9 (b) 1-40 min-ahead prediction with a prediction horizon increment of 4 min. The results are retained using randomly selected data from the testing data set. The red line in the middle box represents the median, the blue box shows the lower quartile and the upper quartile of data distributions, and the black dashed line represents the most extreme values. It is noted that the observation values are within the upper and lower quartiles of the predicted time to apnea onset in Figure 7-9 (a) and that the expected predicted values with the variance covers the expected observation values in Figure 7-9 (b). These results suggest that the predicted values can capture the observations with different

prediction horizons. It is also noted from Figure 7-9 (a) that the prediction errors decrease with the decrease in prediction horizons. We also see this trend in Figure 7-9 (b) in terms of absolute residuals, when the prediction horizon decreases to a 1-5 min-ahead prediction.

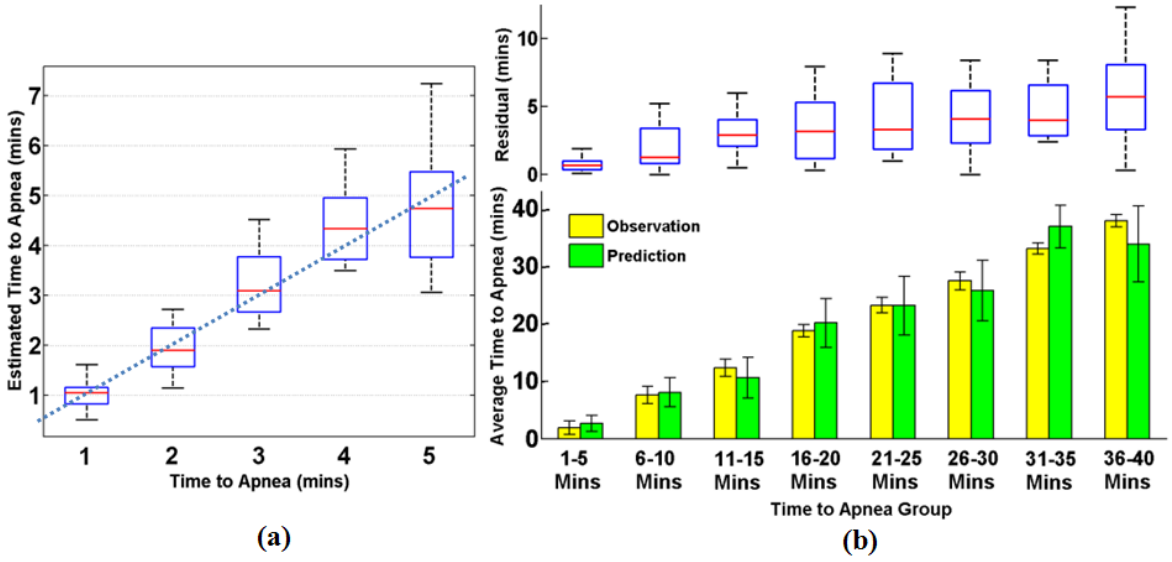


Figure 7-9 Prediction performance of the OSA prognostics

The prognostic performance is further investigated by analyzing the distribution of the estimated risk indicators (see Figure 7-10). The estimated risk indicator is a prognostic quantifier that suggests the likelihood that the failure will happen within a specific time. Figure 7-10 shows the distribution of the risk indicators at 1-5 mins preceding sleep apnea onset. For example, the risk indicator at 3-min to apnea equals the probability of the distribution of time to failure and is evaluated at $t = 3$. The average estimates of risk indicators at 1 to 5 min to apnea onset are 85.8±5.7%, 80.2±10.1%, 75.5±10.8%, 64.9±9.4%, and 60.6±11.1%, respectively. It is noted that the risk indicator is comparably high (over 85% and 5.7 % standard deviation). Furthermore, it is also observed that the estimated risk indicators at 1-3 min till apnea onset are all higher than 75%. Such high prognosis performance can provide reliable supportive information for preventive treatment before the actual apnea happens.

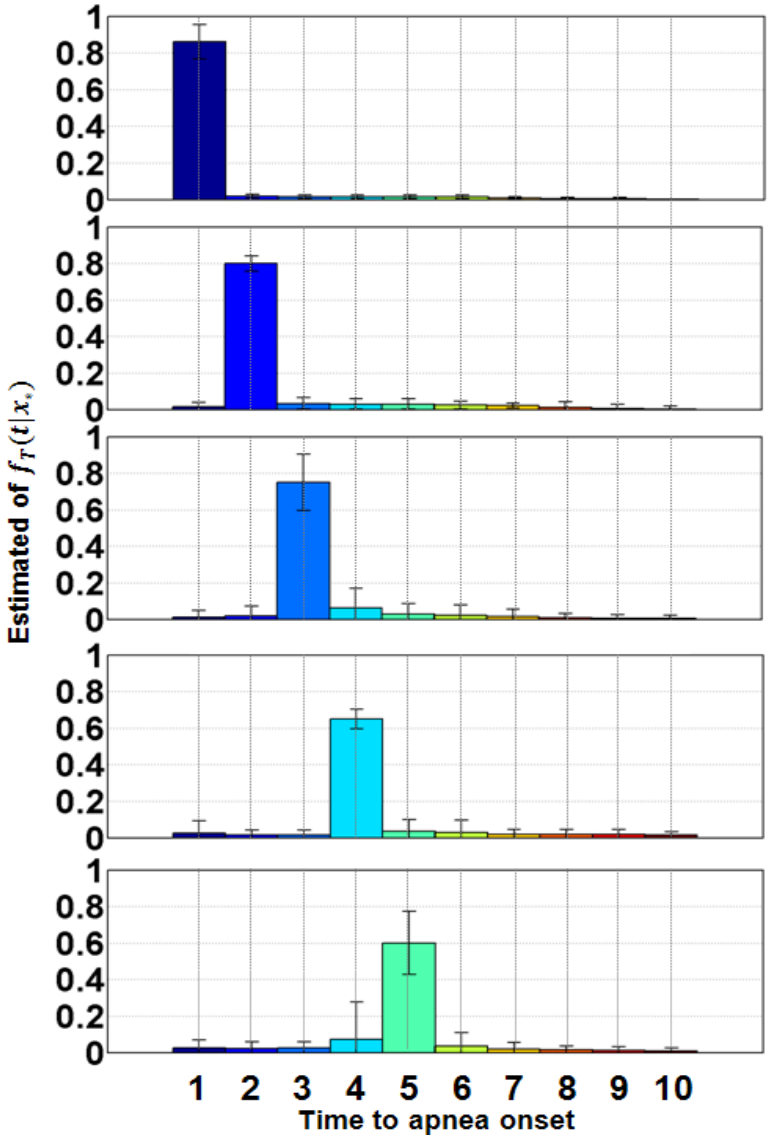


Figure 7-10 Estimated of $f_T(t|\boldsymbol{x}_*)$ for 1-5 min apnea ahead

7.5 Conclusions

We have developed a prognostic method to estimate the distribution of RUL by considering the evolution of normal state vectors to an abnormal state in the state space of the underlying measured signal dynamics. A representation of the state vector transitions as the state space network that is able to optimally separate the state vectors in regions with different dynamic characteristics and visually track the dynamic transitions of the abnormal states in the state space

is presented to facilitate the implementation of the prognosis method. A case study to predict the time to sleep apnea onset has been investigated to validate the prognostic model. The average prediction accuracies (R^2) of the proposed prognostic method are 70-80%. It was also evident that the predicted values can forecast the observations with prediction horizons ranging from 1-40 min ahead. The risk indicators derived from the RUL distribution are 85.8±5.7%, 80.2±10.1%, and 75.5±10.8% at 1 to 3 min to apnea onset, respectively. Such early predictions with reliable risk indicators can be used to support preventive treatment and mitigate the consequences of acute disorders.

Future work focuses on validating the approach with life-threatening acute cardiovascular disorders such as myocardial infarction and ventricular and atrial fibrillations. Early prediction and reliable risk estimators from the proposed method will facilitate the application of this method to real-time tracking of the patient's health status in order to provide timely intervention and preventive treatment.

REFERENCES

[1] *National Center for Chronic Disease Prevention and Health Promotion: Chronic Diseases the Power to Prevent, the Call to Control 2009*. Available: http://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/chronic.pdf

[2] A. R. Schwartz, *et al.*, "Obesity and Obstructive Sleep Apnea: Pathogenic Mechanisms and Therapeutic Approaches," *Proceedings of the American Thoracic Society,* vol. 5, 185-92, Feb 15 2008.

[3] G. D. Foster, *et al.*, "Obstructive Sleep Apnea among Obese Patients with Type 2 Diabetes," *Diabetes Care,* vol. 32, 1017-9, Jun 2009.

[4] A. Romero-Corral, *et al.*, "Interactions between Obesity and Obstructive Sleep Apnea: Implications for Treatment," *Chest,* vol. 137, 711-9, Mar 2010.

[5] *2008–2013 Action Plan for the Global Strategy for the Prevention and Control of Non Communicable Diseases.* : World Health Organization, 2007.

[6] A. N. Vgontzas*, et al.*, "Sleep Apnea and Sleep Disruption in Obese Patients," *Archives of Internal Medicine,* vol. 154*,* 1705, 1994.

[7] A. M. Arruda-Olson*, et al.*, "Sleep Apnea and Cardiovascular Disease," *Herz,* vol. 28*,* 298-303, 2003.

[8] L. Hood*, et al.*, "Systems Biology and New Technologies Enable Predictive and Preventative Medicine," *Science,* vol. 306*,* 640-643, 2004.

[9] W. G. Bradley*, et al.*, "Globalization of P4 Medicine: Predictive, Personalized, Preemptive, and Participatory--Summary of the Proceedings of the Eighth International Symposium of the International Society for Strategic Studies in Radiology," *Radiology,* vol. 258*,* 571-582, 2011.

[10] L. A. Simmons*, et al.*, "Personalized Medicine Is More Than Genomic Medicine: Confusion over Terminology Impedes Progress Towards Personalized Healthcare," *Personalized Medicine,* vol. 9*,* 85-91, Feb 2012.

[11] A. Garratt*, et al.*, "Quality of Life Measurement: Bibliographic Study of Patient Assessed Health Outcome Measures," *British Medical Journal,* vol. 324*,* 1417, 2002.

[12] A. Porta*, et al.*, "Complexity and Nonlinearity in Short-Term Heart Period Variability: Comparison of Methods Based on Local Nonlinear Prediction," *IEEE Transactions on Biomedical Engineering,* vol. 54*,* 94-106, 2007.

[13] V. Tuzcu*, et al.*, "Decrease in the Heart Rate Complexity Prior to the Onset of Atrial Fibrillation," *Europace,* vol. 8*,* 398-402, 2006.

[14] S. Vikman*, et al.*, "Altered Complexity and Correlation Properties of R-R Interval Dynamics before the Spontaneous Onset of Paroxysmal Atrial Fibrillation," *Circulation,* vol. 100*,* 2079-2084, 1999.

[15] U. Pulkkinen, "A Stochastic Model for Wear Prediction through Condition Monitoring," *Operational reliability and systematic maintenance,* 223-243, 1991.

[16] A. Ray and S. Tangirala, "A Nonlinear Stochastic Model of Fatigue Crack Dynamics," *Probabilistic Engineering Mechanics,* vol. 12, 33-40, 1997.

[17] Y. S. Kim and W. J. Kolarik, "Real-Time Conditional Reliability Prediction from on-Line Tool Performance Data," *International Journal of Production Research,* vol. 30, 1831-1844, 1992.

[18] L. Huitian, *et al.*, "Real-Time Performance Reliability Prediction," *IEEE Transactions on Reliability,* vol. 50, 353-357, 2001.

[19] R. B. Chinnam, "On-Line Reliability Estimation for Individual Components Using Statistical Degradation Signal Models," *Quality and Reliability Engineering International,* vol. 18, 53-73, 2002.

[20] L. Jianhui, *et al.*, "Model-Based Prognostic Techniques [Maintenance Applications]," in *IEEE Systems Readiness Technology Conference*, 2003, pp. 330-340.

[21] D. Chelidze and J. P. Cusumano, "A Dynamical Systems Approach to Failure Prognosis," *Journal of Vibration and Acoustics,* vol. 126, 2-8, 2004.

[22] C. J. Lu and W. O. Meeker, "Using Degradation Measures to Estimate a Time-to-Failure Distribution," *Technometrics,* vol. 35, 161-174, 1993.

[23] H. Lu, *et al.*, "Real-Time Performance Reliability Prediction," *IEEE Transactions on Reliability,* vol. 50, 353-357, 2001.

[24] N. Z. Gebraeel, *et al.*, "Residual-Life Distributions from Component Degradation Signals: A Bayesian Approach," *IIE Transactions,* vol. 37, 543-557, 2005.

[25] B. Saha, *et al.*, "Comparison of Prognostic Algorithms for Estimating Remaining Useful Life of Batteries," *Transactions of the Institute of Measurement and Control,* vol. 31, 293-308, 2009.

[26] C. M. Bishop, *Neural Networks for Pattern Recognition*: Oxford University Press, 1995.

[27] N. Gebraeel, *et al.*, "Residual Life Predictions from Vibration-Based Degradation Signals: A Neural Network Approach," *IEEE Transactions on Industrial Electronics,* vol. 51, 694-700, 2004.

[28] K. Goebel, *et al.*, "A Comparison of Three Data-Driven Techniques for Prognostics," in *62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT)*, 2008, pp. 119-131.

[29] R. Polikar, "Ensemble Based Systems in Decision Making," *IEEE Transactions on Circuits and Systems Magazine,* vol. 6, 21-45, 2006.

[30] F. Di Maio, *et al.*, "Ensemble-Approaches for Clustering Health Status of Oil Sand Pumps," *Expert Systems with Applications,* vol. 39, 4847-4859, 2012.

[31] I. Bratko, *et al.*, *Kardio: A Study in Deep and Qualitative Knowledge for Expert Systems*: MIT Press, 1990.

[32] L. J. Shaw, *et al.*, "Prognostic Value of Cardiac Risk Factors and Coronary Artery Calcium Screening for All-Cause Mortality1," *Radiology,* vol. 228, 826-833, 2003.

[33] M. Kikuya, *et al.*, "Prognostic Significance of Blood Pressure and Heart Rate Variabilities the Ohasama Study," *Hypertension,* vol. 36, 901-906, 2000.

[34] T. Ohkubo, *et al.*, "Prognostic Significance of the Nocturnal Decline in Blood Pressure in Individuals with and without High 24-H Blood Pressure: The Ohasama Study," *Journal of hypertension,* vol. 20, 2183-2189, 2002.

[35] R. Hachamovitch, *et al.*, "Comparison of the Short-Term Survival Benefit Associated with Revascularization Compared with Medical Therapy in Patients with No Prior Coronary Artery Disease Undergoing Stress Myocardial Perfusion Single Photon Emission Computed Tomography," *Circulation,* vol. 107, 2900-2907, 2003.

[36] J. P. Halcox, *et al.*, "Prognostic Value of Coronary Vascular Endothelial Dysfunction," *Circulation,* vol. 106, 653-658, 2002.

[37] T. Heitzer, *et al.*, "Endothelial Dysfunction, Oxidative Stress, and Risk of Cardiovascular Events in Patients with Coronary Artery Disease," *Circulation,* vol. 104, 2673-2678, 2001.

[38] W. E. Hellings, *et al.*, "Composition of Carotid Atherosclerotic Plaque Is Associated with Cardiovascular Outcome a Prognostic Study," *Circulation,* vol. 121, 1941-1950, 2010.

[39] P. Verdecchia, *et al.*, "Ambulatory Pulse Pressure a Potent Predictor of Total Cardiovascular Risk in Hypertension," *Hypertension,* vol. 32, 983-988, 1998.

[40] S. Okuno, *et al.*, "Presence of Abdominal Aortic Calcification Is Significantly Associated with All-Cause and Cardiovascular Mortality in Maintenance Hemodialysis Patients," *American Journal of Kidney Diseases,* vol. 49, 417, 2007.

[41] P. Verdecchia, *et al.*, "Prognostic Significance of Blood Pressure Variability in Essential Hypertension," *Blood Pressure Monitoring,* vol. 1, 3, 1996.

[42] P. W. F. Wilson, *et al.*, "Prediction of Coronary Heart Disease Using Risk Factor Categories," *Circulation,* vol. 97, 1837-1847, 1998.

[43] W. C. Levy, *et al.*, "The Seattle Heart Failure Model," *Circulation,* vol. 113, 1424-1433, 2006.

[44] D. A. Morrow, "Cardiovascular Risk Prediction in Patients with Stable and Unstable Coronary Heart Disease," *Circulation,* vol. 121, 2681-2691, 2010.

[45] R. L. Benza, *et al.*, "Predicting Survival in Pulmonary Arterial Hypertension," *Circulation,* vol. 122, 164-172, 2010.

[46] A. R. Sharrett, *et al.*, "Coronary Heart Disease Prediction from Lipoprotein Cholesterol Levels, Triglycerides, Lipoprotein(a), Apolipoproteins a-I and B, and Hdl Density Subfractions," *Circulation,* vol. 104, 1108-1113, 2001.

[47] I. Shai, *et al.*, "Multivariate Assessment of Lipid Parameters as Predictors of Coronary Heart Disease among Postmenopausal Women," *Circulation,* vol. 110, 2824-2830, 2004.

[48] T. Pischon, *et al.*, "Non High Density Lipoprotein Cholesterol and Apolipoprotein B in the Prediction of Coronary Heart Disease in Men," *Circulation,* vol. 112, 3375-3383, 2005.

[49] S. C. Savva, *et al.*, "Waist Circumference and Waist-to-Height Ratio Are Better Predictors of Cardiovascular Disease Risk Factors in Children Than Body Mass Index," *International Journal of Obesity and Related Metabolic Disorders,* vol. 24, 1453-8, 2000.

[50] M. J. Pencina, *et al.*, "Predicting the 30-Year Risk of Cardiovascular Disease," *Circulation,* vol. 119, 3078-3084, 2009.

[51] N. Allen, *et al.*, "Impact of Blood Pressure and Blood Pressure Change During Middle Age on the Remaining Lifetime Risk for Cardiovascular Disease / Clinical Perspective," *Circulation,* vol. 125, 37-44, 2012.

[52] P. M. Ridker, *et al.*, "Comparison of C-Reactive Protein and Low-Density Lipoprotein Cholesterol Levels in the Prediction of First Cardiovascular Events," *New England Journal of Medicine,* vol. 347, 1557-1565, 2002.

[53] J. J. Cao, *et al.*, "Association of Carotid Artery Intima-Media Thickness, Plaques, and C-Reactive Protein with Future Cardiovascular Disease and All-Cause Mortality," *Circulation,* vol. 116, 32-38, 2007.

[54] P. M. Ridker, *et al.*, "C-Reactive Protein and Parental History Improve Global Cardiovascular Risk Prediction," *Circulation,* vol. 118, 2243-2251, 2008.

[55] J. T. Saunders, *et al.*, "Cardiac Troponin T Measured by a Highly Sensitive Assay Predicts Coronary Heart Disease, Heart Failure, and Mortality in the Atherosclerosis Risk in Communities Study / Clinical Perspective," *Circulation,* vol. 123, 1367-1376, 2011.

[56] L. J. Shaw, *et al.*, "Use of a Prognostic Treadmill Score in Identifying Diagnostic Coronary Disease Subgroups," *Circulation,* vol. 98, 1622-1630, 1998.

[57] P. M. Okin, *et al.*, "Assessment of Qt Interval and Qt Dispersion for Prediction of All-Cause and Cardiovascular Mortality in American Indians : The Strong Heart Study," *Circulation,* vol. 101, 61-66, 2000.

[58] J. A. Gomes, *et al.*, "Prediction of Long Term Outcomes by Signal Averaged Electrocardiography in Patients with Unsustained Ventricular Tachycardia, Coronary Artery Disease, and Left Ventricular Dysfunction," *Circulation,* vol. 104, 436-441, 2001.

[59] R. S. Crow, *et al.*, "Prognostic Significance of Corrected Qt and Corrected Jt Interval for Incident Coronary Heart Disease in a General Population Sample Stratified by Presence or Absence of Wide Qrs Complex," *Circulation,* vol. 108, 1985-1989, 2003.

[60] L. F. Bielak, *et al.*, "Probabilistic Model for Prediction of Angiographically Defined Obstructive Coronary Artery Disease Using Electron Beam Computed Tomography Calcium Score Strata," *Circulation,* vol. 102, 380-385, 2000.

[61] R. Vliegenthart, *et al.*, "Coronary Calcification Improves Cardiovascular Risk Prediction in the Elderly," *Circulation,* vol. 112, 572-577, 2005.

[62] K. Steel, *et al.*, "Complementary Prognostic Values of Stress Myocardial Perfusion and Late Gadolinium Enhancement Imaging by Cardiac Magnetic Resonance in Patients with Known or Suspected Coronary Artery Disease," *Circulation,* vol. 120, 1390-1400, 2009.

[63] A. Dortimer, *et al.*, "Distribution of Coronary Artery Disease. Prediction by Echocardiography," *Circulation,* vol. 54, 724-729, 1976.

[64] K. Lee, *et al.*, "Prognostic Value of Radionuclide Angiography in Medically Treated Patients with Coronary Artery Disease. A Comparison with Clinical and Catheterization Variables," *Circulation,* vol. 82, 1705-1717, 1990.

[65] M. J. Budoff, *et al.*, "Continuous Probabilistic Prediction of Angiographically Significant Coronary Artery Disease Using Electron Beam Tomography," *Circulation,* vol. 105, 1791-1796, 2002.

[66] N. P. Paynter, *et al.*, "Cardiovascular Disease Risk Prediction with and without Knowledge of Genetic Variation at Chromosome 9p21.3," *Annals of Internal Medicine,* vol. 150, 65-72, 2009.

[67] S. E. Humphries, *et al.*, "Coronary Heart Disease Risk Prediction in the Era of Genome-Wide Association Studies," *Circulation,* vol. 121, 2235-2248, 2010.

[68] R. S. Khattar, *et al.*, "Prediction of Coronary and Cerebrovascular Morbidity and Mortality by Direct Continuous Ambulatory Blood Pressure Monitoring in Essential Hypertension," *Circulation,* vol. 100, 1071-1076, 1999.

[69] T. Le, *et al.*, "Wireless Wearable Multisensory Suite and Real-Time Prediction of Sleep Apnea Episodes," *IEEE Transactions on Translational Engineering in Health and Medicine,* 2013 (In Press).

[70] M. Kaufmann and D. Wagner, *Drawing Graphs: Methods and Models* vol. 2025: Springer Verlag, 2001.

[71] M. D. Escobar and M. West, "Bayesian Density Estimation and Inference Using Mixtures," *American Statistical Association,* vol. 90, 577-588, 1995.

[72] A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," *Statistics and Computing,* vol. 14, 199-222, 2004.

[73] F. Takens, "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, ed: Springer, 1981, pp. 366-381.

[74] A. M. Fraser and H. L. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information," *Physical Review A,* vol. 33, 1134, 1986.

[75] S. Boccaletti, *et al.*, "Reconstruction Embedding Spaces of Coupled Dynamical Systems from Multivariate Data," *Physical Review E,* vol. 65, 035-204, 2002.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

In summary, this research provides mathematical and computational schemes to address diagnostic and prognostic issues in realizing the proactive and personalized diagnosis and treatment of cardiovascular disorders. Specifically, attention has focused on four complementary aspects of the development of a comprehensive P4 system of healthcare for the treatment of cardiovascular disorders: (1) a quantitative model that can capture the underlying cardiorespiratory couplings and generate noninvasive surrogate hemodynamic signals, (2) high-specificity diagnostic methods to identify and localize disorders, (3) real-time prediction methods that can drive advanced prognostic and preventive therapies, and (4) prognostic approaches that provide accurate risk indicators and survival assessments related to the disease's progression. The research facilitates a shift in the healthcare paradigm from in-hospital to out-of-hospital (at-home) diagnostics, and personalized, proactive healthcare, especially for chronically ill and at-risk-patients, who make up $> 40\%$ of the US population and account for $\sim 75\%$ of current healthcare costs. The major conclusions and future work are as follows.

8.1 Conclusions

A quantitative data-driven model of the cardiovascular system was developed. ECG features were used to construct atrioventricular excitation inputs for a nonlinear deterministic lumped parameter model of cardiovascular system dynamics. The model is able to generate multiple

synchronized hemodynamic signals from cardiovascular systems in real time. The model provided an efficient tool for quantitatively assessing the underlying couplings between the mechanical and electrical components of heart dynamics and an ECG signal. The surrogate hemodynamic signals generated from the model outputs can lead to the development of responsive and cost-effective medical devices especially for point-of care diagnosis and personalized treatment.

A stochastic representation of the complex cardiac excitation and propagation dynamics as a random walk on a network reconstructed from Vectorcardiogram (VCG) signals was investigated for the detection and localization of cardiovascular disorders. Quantifiers extracted from the random walk network were used to characterize the nonlinear spatiotemporal dynamics of cardiac vectors and the recurrent patterns of Vectorcardiograms (VCGs). Extensive tests with signals from the PTB database of the PhysioNet databank suggest that locations of MI can be determined accurately (sensitivity of ~88% and specificity of ~92%) from tracking certain consistently estimated invariants of this random walk representation.

A novel feature extraction method and a prediction model were introduced to track and forecast the evolution of the coupled nonlinear and nonstationary cardiorespiratory dynamics underlying the measured physiological signals. Nonlinear and nonstationary characteristics of the features were captured by using power spectrum and recurrence quantification analysis. A Dirichlet process-based Gaussian process was utilized to forecast the complex dynamic evolution of the extracted signatures. A case study to predict the onset of sleep apnea episodes with the signals collected from the multisensory suite developed at COMMSENS Okstate Lab as well as PhysioNet's OSA database suggests that the average accuracy for predicting an OSA episode 1-min ahead is 83% and 3-min ahead is 77%.

An advanced prognostic model to derive the remaining useful life distribution of cardiovascular disorders was investigated. The prognostic model used the stochastic evolution of a transition pathway from a normal state to an abnormal state vector in the clustered state space using the Dirichlet process to estimate the remaining useful time distribution and risk indicators. The implementation of the prognostic model was facilitated by introducing the transition state space network, which was able to visually represent the transition pathways and optimally separate the state vectors in clusters with different dynamic characteristics. A case study using the prognostic model to derive the remaining time to sleep apnea onset distribution was investigated and the estimated risk indicators were reported as 85.8±5.7%, 80.2±10.1%, 75.5±10.8% at 1 to 3 min till apnea onset, respectively. Such early predictions with reliable risk indicators can be used to drive responsive preventive treatments before an actual apnea episode occurs.

8.2 Future work

Future research emerging from the current work will focus on the following:

First, the prognostic approach will be further developed for real-time estimation of remaining useful time distributions toward life-threatening cardiovascular disorders such as myocardial infarction and ventricular fibrillation. The complex short-term transitions in the dynamics of these life-threatening acute disorders need to be fully addressed in order to achieve an efficient prognostic model. Furthermore, since a prediction type II error (false negative) in these types of diseases might cause serious consequences, the prognostic methods also need to provide a reliable risk indicator that physicians can use for timely interventions before deadly events happen.

Second, the preventive noninvasive treatment of cardiorespiratory disorders such as sleep apnea will be investigated to take advantage of the prediction outcomes. Preventive noninvasive

treatments, from the engineering point of view, will provide adaptive control algorithms for the medical devices currently used for treating these disorders. Furthermore, we are also investigating new intervention treatment methods using the comparably high, accurate, and timely prediction results from the prognostic model to mitigate or invert the degradation process of acute or chronic disorders before they actually occur.

Third, ongoing investigations are focused on the development of instrumentation that can advance the data collecting processes for real-time implementation of the prognostic model. The current monitoring system requires physical measurements of physiological signals from the sensors. Although advancements in wireless and information technology provide high-quality data, high spatiotemporal resolution and artifact-free signals are still the goals. Magnetoencephalography (MEG) and magnetocardiology (MCG) are new technologies that record magnetic fields for mapping the electrical activities with the physical functions of the brain and heart. Investigation of magnetic field instruments that can wirelessly capture spatiotemporal dynamics and functions of the heart and brain will facilitate the online implementation of the prognostic model.

APPENDIX

These following Matlab toolboxes have been used for the implementations of the methodology:

- Recurrence Quantification Analysis: http://tocsy.pik-potsdam.de/CRPtoolbox/
- Gaussian Process: http://www.gaussianprocess.org/gpml/code/matlab/doc/
- Network analysis: http://strategic.mit.edu/downloads.php?page=matlab_networks

The remainder of the appendix is organized as follows:

A1. Simulink model

A2. Atrioventricular activation function

A3. VCG random walk network

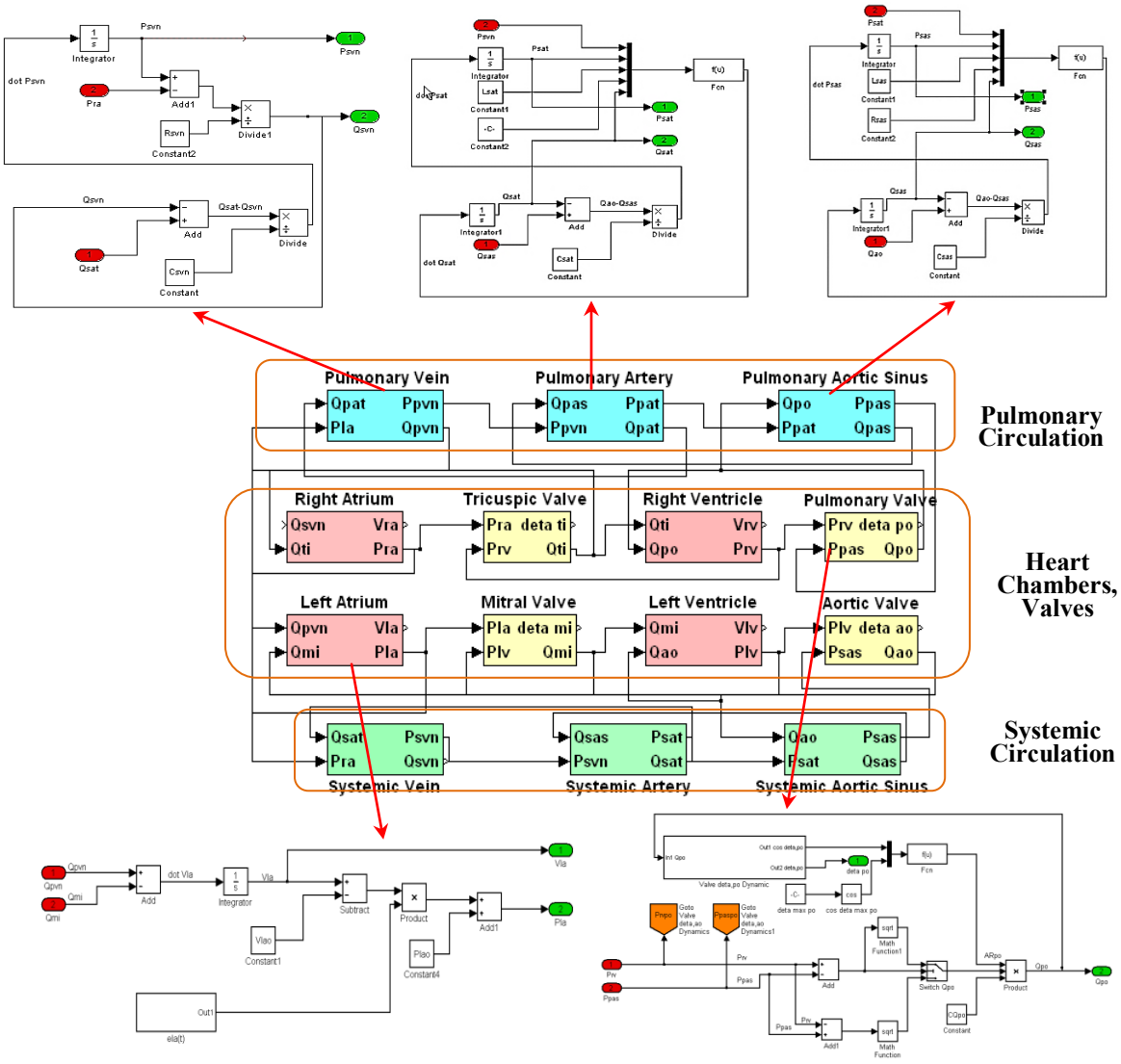
A4. CART classification

A5. Dirichlet process based Gaussian process mixture (DPMG) prediction

A6. Color coded state space network representation

A7. Estimation of time to failure distribution

A.1 Simulink model



Simulink implementation of cardiovascular model

A.2 Atrioventricular activation function

```
%% Initialize the parameters

Elamax = 0.2500;
Elamin = 0.1500;
Elvd =0.1000;
Elvs =2.5000;
Eramax =0.2500;
Eramin =0.1500;
```

```
Ervd =0.1000;
Ervs =1.1500;
PatOutGain_opt =[];
PatOutOff_opt =[];
ResOutGain_opt =[];
ResOutOff_opt =[];
artOutGain_opt =[];
artOutOff_opt =[];
a=5.5;
b=5.5;

%% Load the database

DataSource = 'L:\Research\CurrentModel\DataBase\GoodOneMinute.txt';
strfilename = importdata(DataSource);
for k =1:1:length(strfilename)
    names(k,:) = strfilename{k,1};
end
NumOfRec = length(strfilename);       % Number of record to read
Records = [1:35];
Records([4 9 10 11 12 14 18 19 20 23 26 32]) = [];
countfea = 0;
for index = [1 2 5 6 7 13 15 17 21 27 28 29 31 36:42]
    countfea =countfea + 1;
     index = 2
    data = [];
    dataResp = [];
    IndexStr = char(strcat(strfilename(index)))
    dataload =
load(char(strcat('L:\Research\CurrentModel\DataBase\GoodOneMinute\',Ind
exStr,'m.mat')));

%%Header reading

    heastr =
char(strcat('L:\Research\CurrentModel\DataBase\GoodOneMinute\',IndexStr
,'.hea'));
    f = fopen(heastr,'r');
    z = fgetl(f);
    B = sscanf(z, '%*s %d %d %d',[1,3]);
    nosig = B(1);  % number of signals
    nosig = nosig ; % .dat files
    Sfreq =B(2);   % sample rate of data
    clear A;
    for k=1:nosig
        z = fgetl(f);
        A= sscanf(z, '%*s %f %s',[1 2]);
        %detect the format of the hearder with/without offset and unit
        if strfind(char(A(2:end)),')/m') %with offset with unit
            A = sscanf(z, '%*s %d %f %*c %f %*s',[1 3]);
        elseif  strfind(char(A(2:end)),'/m') %without offset with unit
            A = sscanf(z, '%*s %d %f',[1 2]);
            A(3) = 0;
```

```
        else
            A = sscanf(z, '%*s %d %f',[1 2]); %without offset without
unit
            A(3) = 0;
        end
        dformat(k)= A(1);           % format; here only 212 is allowed
        gain(k)= A(2);              % number of integers per mV
        base(k) = A(3);             % offset value
        clear A;
    end;
    fclose(f);
    data1 = dataload.val;
    samp_freq = Sfreq ;

%% Initialize every time run the simulink optimization

    PatOutOff = 0; %initialize gain of Resp out for the optimization
design
    PatOutGain = 1;
    ResOutGain = 1/6;
    ResOutOff = -1;
    artOutOff= 0;
    artOutGain= 1;

    Elamax = 0.2500;
    Elamin = 0.1500;
    Elvd =0.1000;
    Elvs =2.5000;
    Eramax =0.2500;
    Eramin =0.1500;
    Ervd =0.1000;
    Ervs =1.1500;

%% Detection of PQRST

    D = data;
    SampFreq =samp_freq;
    windowSize = 5;
    filsig = filter (ones(1,windowSize)/windowSize,1,D);
    y = medfilt1(filsig,0.2*SampFreq); % 1st median filter
    s1 = y;
    y = medfilt1(s1,0.6*SampFreq); % 2nd median filter
    NoWander= filsig - y;
    D = transpose (NoWander);
    D = cwt (D, 1:4, 'bior2.4');
    D = transpose (D);
    Wave4 = D (:,4);
    flagP = 0;
    flagT = 0;
    Ppeaks = [];
    Qpeaks = [];
    Rpeaks = [];
    Speaks = [];
```

```
Tpeaks = [];
PRSegment = [];
QRSegment = [];
RSSegment = [];
RTSegment = [];
PRSeg = [];
STSeg = [];
P =[];
Q =[];
R =[];
S =[];
T =[];
Anno = [];

[Rpeaks,HeartRate,R_series,RR] = rpeak1(NoWander,samp_freq);
%Adaptive value
RRAvgTemp = mean(diff(Rpeaks));
% RRAvg = mean(R_series)
PR = RRAvgTemp/3;
KQ = RRAvgTemp/25;
QR = RRAvgTemp/15;
RS = RRAvgTemp/4;
SJ = RRAvgTemp/25;
RT = RRAvgTemp/1.75;
ToT = RRAvgTemp/7;

WindowPR = fix(PR);
WindowQR = fix(QR);
WindowRS = fix(RS);
WindowRT = fix(RT);
WindowKQ = fix(KQ);
WindowSJ =  fix(SJ);
WindowToT = fix(RRAvgTemp/6);
WindowPPof = fix(RRAvgTemp/7);

for step =1:length(Rpeaks)
    SeriesP = 0;
    SeriesQ = 0;
    SeriesS = 0;
    SeriesT = 0;
    flagP = 0;  %autofill the P if not detected
    flagT = 0;
    %Detect if the ECG start with R but not have P,Q,S,T
    if (Rpeaks(step)-WindowPR)> 0
        PRSegment = NoWander((Rpeaks(step)-WindowPR):Rpeaks(step));
    else
        break;
    end
    if (Rpeaks(step)-WindowQR) > 0
        QRSegment = Wave4((Rpeaks(step)-WindowQR):Rpeaks(step));
    else
        break;
    end
```

```
        if (Rpeaks(step)+WindowRS)< length(Wave4)
%dectec if exceed length signal-> no S after last R peaks)
            RSSegment = Wave4(Rpeaks(step):(Rpeaks(step)+WindowRS));
        else
            break;
        end
        if (Rpeaks(step)+WindowRT)< length(NoWander)
%dectec if exceed length signal-> no T after last R peaks)
            RTSegment = NoWander(Rpeaks(step):(Rpeaks(step)+WindowRT));
        else
            break;
        end
        %P Detection

        for Pinterval = 2:WindowPR
            if PRSegment(Pinterval)> PRSegment(Pinterval-1) && ...
                    PRSegment(Pinterval)> PRSegment(Pinterval+1)&&...
                    PRSegment(Pinterval)> SeriesP
                SeriesP = PRSegment(Pinterval);
                Ppoint = Pinterval-1;
                flagP = 1; %detect P
            end
        end
        if flagP == 1
            Ppeaks(step) = Rpeaks(step) - WindowPR + Ppoint;
        else
            Ppeaks(step) = Rpeaks(step) - 0.2*SampFreq;
        end
        %P off
        Pofpeaks(step) = Ppeaks(step) + WindowPPof;
        %     Pof_series(step) = NoWander(Pofpeaks(step));
        %Q Detection
        count =1;
        for Qinterval = 2:WindowQR
            if QRSegment(Qinterval)< QRSegment(Qinterval-1) && ...
                    QRSegment(Qinterval)< QRSegment(Qinterval+1)&&...
                    QRSegment(Qinterval)< 0
                SeriesQ(count) = QRSegment(Qinterval);
                Qpoint_temp(count) = Qinterval-1;
                count = count +1;
            end
        end
        [MinQVal,Qposi] = min(SeriesQ);
        Qpoint = Qpoint_temp(Qposi);
        Qpeaks(step) = Rpeaks(step) - WindowQR + Qpoint;
        %K Detection
        KQSegment = Wave4((Qpeaks(step)-WindowQR):Qpeaks(step));
        count = 1;
        for Sinterval = 2:WindowRS
            if RSSegment(Sinterval)< RSSegment(Sinterval-1) && ...
                    RSSegment(Sinterval)< RSSegment(Sinterval+1)&&...
                    RSSegment(Sinterval)< 0
                SeriesS(count) = RSSegment(Sinterval);
```

```
                Spoint_temp(count) = Sinterval -1;
                count = count +1;
            end
        end
        [MinSVal,Sposi] = min(SeriesS);
        Spoint = Spoint_temp(Sposi);   %To find the closest of minima S
to R
        Speaks(step) = Rpeaks(step) + Spoint;
        %J Detection
        SJSegment = Wave4(Speaks(step):(Speaks(step)+WindowSJ));
        for Tinterval = 2:WindowRT
            if RTSegment(Tinterval)> RTSegment(Tinterval-1) && ...
                    RTSegment(Tinterval)> RTSegment(Tinterval+1)&&...
                    RTSegment(Tinterval)> SeriesT
                SeriesT = RTSegment(Tinterval);
                Tpoint = Tinterval-1;
                flagT =1;
            end
        end
        if flagT == 1
            Tpeaks(step) = Rpeaks(step) + Tpoint;
        else
            Tpeaks(step) = Rpeaks(step) + 0.25*SampFreq;
        end
        %T onset Detection
        Topeaks(step) = Tpeaks(step) - WindowToT;
        %     To_series(step) = NoWander(Topeaks(step));
        %
        %         Ppeaks(step) = Rpeaks(step) - WindowPR + Ppoint;
        P_series(step) =  SeriesP;
        %     Kpeaks(step) = Qpeaks(step) - WindowKQ + Kpoint;
        %     K_series(step) =  NoWander(Kpeaks(step));
        Qpeaks(step) = Rpeaks(step) - WindowQR + Qpoint;
        Q_series(step) =  NoWander(Qpeaks(step));
        Speaks(step) = Rpeaks(step) + Spoint +1;
        S_series(step) =   NoWander(Speaks(step));
        %     Jpeaks(step) = Speaks(step) + Jpoint;
        %     J_series(step) =  NoWander(Jpeaks(step));
        %         Tpeaks(step) = Rpeaks(step) + Tpoint;
        T_series(step) =  SeriesT;
    end
    Rpeaks = Rpeaks(1:length(Ppeaks));

%% Create the activation signal
    Activ = zeros(1,length(data)); %Inintial Activation value;
    ActivVent = zeros(1,length(data)); %Inintial Activation value for
Ventricles;
    ActivAtr = zeros(1,length(data));  %%Inintial Activation value for
Artria

    for BeatOrder= 1:1:length(Rpeaks)     %position of the R peaks
```

```
        AtrAct =
ECGActivation_Atrium(Ppeaks(BeatOrder),Rpeaks(BeatOrder));
        VentAct =
ECG_Vent(Rpeaks(BeatOrder),Tpeaks(BeatOrder),Tpeaks(BeatOrder);
        ActivVent(1,Rpeaks(BeatOrder):Tpeaks(BeatOrder))= VentAct;
        ActivAtr(1,Ppeaks(BeatOrder):Rpeaks(BeatOrder)) = AtrAct;
    end

    time1= 1/samp_freq:1/samp_freq:length(data)/samp_freq;
    ArtriActive = [];
    VentriActive = [];
    a=0;
    b=5.5;
    ArtriActive(:,1) = time1;
    ArtriActive(:,2) = ActivAtr';
    Resp_Art(:,1) = time1;
    Resp_Art(:,2) = dataResp';
    % ArtriActive(:,2) =  dataResp/10;
    VentriActive(:,1) = time1;
    VentriActive(:,2) = ActivVent;
    Resp_Vent(:,1) = time1;
    Resp_Vent(:,2) =  dataResp;
    HeartParameters;    Samp_Time = 1/samp_freq ;%Initialize parameters
for
    ParameterLong =ParameterLong';
    PatOutGain= ParameterLong(1,countfea) ;
    PatOutOff= ParameterLong(2,countfea) ;
    ResOutGain= ParameterLong(3,countfea) ;
    ResOutOff= ParameterLong(4,countfea) ;
    artOutGain= ParameterLong(5,countfea) ;
    artOutOff= ParameterLong(6,countfea) ;
    a= ParameterLong(7,countfea) ;
    b= ParameterLong(8,countfea) ;
    Elvs= ParameterLong(9,countfea)  ;
    Ervs= ParameterLong(10,countfea) ;
    Elvd= ParameterLong(11,countfea) ;
    Ervd= ParameterLong(12,countfea)  ;
    DetaMaxAo= ParameterLong(13,countfea) ;
    DetaMaxPo= ParameterLong(14,countfea) ;
    Elamax= ParameterLong(15,countfea) ;
    Eramax= ParameterLong(16,countfea)  ;
    DetaMaxMi= ParameterLong(17,countfea)  ;
    DetaMaxTi= ParameterLong(18,countfea);
    Elamin= ParameterLong(19,countfea) ;
    Eramin= ParameterLong(20,countfea) ;
    sim('L:\Research\CurrentModel\CombinationWithECG.mdl',60);
    CorrRes(countfea)=Corr_Resp(1,2);
    CorPat(countfea)=Corr_Pat(1,2);
    CorCVP(countfea)=Corr_CVP(1,2);
    % Filter out the signal
    [ ResModel ] = Fourier (Ppvn.signals.values(2000:end),Sfreq,0,1.7);
    [ PatModel ] = Fourier (Pla.signals.values(2000:end),Sfreq,0,10);
    [ PatModel1 ] = Fourier (Ppul.signals.values(2000:end),Sfreq,0,60);
```

```
    [ CVPModel ] = Fourier (Pra.signals.values(2000:end),Sfreq,0,10);
    MeasRes = dataResp_1(5000:20000)';
    ModelRes = Ppvn.signals.values(5000:20000)*ResOutGain+ResOutOff;
    [MaxVal,Indx] = max(xcorr(ModelRes,MeasRes));
    DelayRes = Indx -length(ModelRes);
    if abs(DelayRes) >1000
        DelayRes = 360; %Prevent too long delay
    end
    Rsq_Resp(countfea) = 1- sum((MeasRes-ModelRes).^2)/sum((MeasRes-
mean(MeasRes)).^2);
    Rsq_Pat(countfea) =1- sum((MeasPat-ModelPat).^2)/sum((MeasPat-
mean(MeasPat)).^2);
    Rsq_CVP(countfea) =1- sum((MeasCVP-ModelCVP).^2)/sum((MeasCVP-
mean(MeasCVP)).^2);
    Corr_Resp1 = corrcoef(MeasRes,ModelRes);
    Corr_Pat1 = corrcoef(MeasPat,ModelPat);
    Corr_CVP1 = corrcoef(MeasCVP,ModelCVP);
    CorrRes1(countfea)=Corr_Resp1(1,2);
    CorPat1(countfea)=Corr_Pat1(1,2);
    CorCVP1(countfea)=Corr_CVP1(1,2);
    samp_freq =360;t1=1/samp_freq;
    set(gca,'fontweight','bold','fontsize',30);
```

A.3 VCG random walk network

```
load('L:\Research\MI Localization\Matlab Octant
Program\Group\CurrentMILocation_Index.mat') % get the MI location index
load('L:\Research\MI Localization\Matlab Octant
Program\Group\FormerMILocation_Index.mat')  % get the former MI index
Infero2 =union(Infero,Infero1);
Infero_Lateral2 = union(Infero_Lateral,Infero_Lateral1);
Infero_Posterior2 = union(Infero_Posterior,Infero_Posterior1);
Infero_Postero_Lateral2 =
union(Infero_Postero_Lateral,Infero_Postero_Lateral1);
Anterior2 = union(Anterior,Anterior1);
Anterio_Septal2 = union(Anterio_Septal,Anterio_Septal1);
Antero_Lateral2 = union(Antero_Lateral,Antero_Lateral1);
Antero_Septo_Lateral2 =
union(Antero_Septo_Lateral,Antero_Septo_Lateral1);
Lateral2= union(Lateral,Lateral1);
Posterior2= union(Posterior,Posterior1);
Postero_Lateral2 = union(Postero_Lateral,Postero_Lateral1);

%% All infarction without former MI

Infero3 =intersect(Infero,MI_NoPosition1);
Infero_Lateral3 = intersect(Infero_Lateral,MI_NoPosition1);
Infero_Posterior3 = intersect(Infero_Posterior,MI_NoPosition1);
Infero_Postero_Lateral3 =
intersect(Infero_Postero_Lateral,MI_NoPosition1);
Anterior3 = intersect(Anterior,MI_NoPosition1);
Anterio_Septal3 = intersect(Anterio_Septal,MI_NoPosition1);
```

```
Antero_Lateral3 = intersect(Antero_Lateral,MI_NoPosition1);
Antero_Septo_Lateral3 = intersect(Antero_Septo_Lateral,MI_NoPosition1);
Lateral3= intersect(Lateral,MI_NoPosition1);
Posterior3= intersect(Posterior,MI_NoPosition1);
Postero_Lateral3 = intersect(Postero_Lateral,MI_NoPosition1);
 A
={Infero3,Infero_Lateral3,Infero_Posterior3,Infero_Postero_Lateral3,...
    Anterior3,Anterio_Septal3,Antero_Lateral3,...
    Lateral3,Posterior3,Postero_Lateral3,HC};
class_name = {'Infero','Infero Lateral','Infero Posterior','Infero
Postero Lateral',...
    'Anterior','Anterio Septal','Antero Lateral',...
    'Lateral','Posterior','Postero Lateral','HC'};
quad_stat = {'Quad0Num' 'Quad0MaxN' 'Quad0AvgN' 'Quad0VarN' ...
    'Quad1Num' 'Quad1MaxN' 'Quad1AvgN' 'Quad1VarN' ...
    'Quad2Num' 'Quad2MaxN' 'Quad2AvgN' 'Quad2VarN' ...
    'Quad3Num' 'Quad3MaxN' 'Quad3AvgN' 'Quad3VarN' ...
    'Quad4Num' 'Quad4MaxN' 'Quad4AvgN' 'Quad4VarN' ...
    'Quad5Num' 'Quad5MaxN' 'Quad5AvgN' 'Quad5VarN' ...
    'Quad6Num' 'Quad6MaxN' 'Quad6AvgN' 'Quad6VarN' ...
    'Quad7Num' 'Quad7MaxN' 'Quad7AvgN' 'Quad7VarN'};

%% Formulate the octant

idxcount = 0;
save_loc =1;
db_idx = 1;
for group = 1:1:length(A)
%   for group =8:9
    group
    adjMat =[]; % Initial the adjacent matrix of each group
    idxcount= 0;    %Start index for each group
    for loop = A{group}'%run the loop for each MI class
        %   for loop = 1
        idxcount = idxcount+1 %count the number of the opened record
        [ecg,IndexStr,H] = ptbopenfile(loop);
        EcgChan = 2; % select ecg channel 2;
        lengthData = 11000; % length of analized data;
        fs = 1000; %sampling frequency
        if strcmp(lower(H.Diagnosis),'healthy control')+
strcmp(lower(H.Diagnosis),'myocardial infarction')==0
            continue;
        end
        for i=13:15 %filtering the VCG signal only
            temp = fftfilter(ecg(:,i),2,40);%bandpass filter
            fs = 1000;
            temp = BaseLine2(temp,0.3*fs,0.6*fs,'mn');
            [baseline, corrected, coeffs] = parabolicfilter(temp);
            EKG(:,i) = temp(500:end-500);
        end
        VCG = EKG(1:lengthData,13:15);
        PeakIdx = nqrsdetect(VCG(:,1),fs); %Identify QRS using filter
bank
```

```
        R_series1 = VCG(PeakIdx,1)' ;
        rr_series = interp1([1 PeakIdx' lengthData],[R_series1(1)
R_series1 R_series1(end)],1:1:lengthData,'nearest'); %formulate the
trainpulse of the R amplitude
        RR1 = diff(PeakIdx);                            %RR interval
        avrRR= mean(RR1)/1000;
        %% Create the octant
        clear EKG;
        clear ecg;
        VQuad7 = [];    VQuad6 = [];
        VQuad5 = [];    VQuad4 = [];
        VQuad3 = [];    VQuad2 = [];
        VQuad1 = [];    VQuad0 = [];
        for i = 1:max(size(VCG))
            if VCG(i,1)<0 & VCG(i,2)<0 & VCG(i,3)<0%---000
                VQuad0 = vertcat(VQuad0,VCG(i,:));
            end
            if VCG(i,1)<0 & VCG(i,2)<0 & VCG(i,3)>0%--+001
                VQuad1 = vertcat(VQuad1,VCG(i,:));
            end
            if VCG(i,1)<0 & VCG(i,2)>0 & VCG(i,3)<0%-+-010
                VQuad2 = vertcat(VQuad2,VCG(i,:));
            end
            if VCG(i,1)<0 & VCG(i,2)>0 & VCG(i,3)>0%-++011
                VQuad3 = vertcat(VQuad3,VCG(i,:));
            end
            if VCG(i,1)>0 & VCG(i,2)<0 & VCG(i,3)<0%+--100
                VQuad4 = vertcat(VQuad4,VCG(i,:));
            end
            if VCG(i,1)>0 & VCG(i,2)<0 & VCG(i,3)>0%+-+101
                VQuad5 = vertcat(VQuad5,VCG(i,:));
            end
            if VCG(i,1)>0 & VCG(i,2)>0 & VCG(i,3)<0%++-110
                VQuad6 = vertcat(VQuad6,VCG(i,:));
            end
            if VCG(i,1)>0 & VCG(i,2)>0 & VCG(i,3)>0%+++111
                VQuad7 = vertcat(VQuad7,VCG(i,:));
            end

        end

        %Length condition
        if (isempty(VQuad0)),VQuad0=[0 0 0];end ;
        if (isempty(VQuad1)),VQuad1=[0 0 0];end ;
        if (isempty(VQuad2)),VQuad2=[0 0 0];end ;
        if (isempty(VQuad3)),VQuad3=[0 0 0];end ;
        if (isempty(VQuad4)),VQuad4=[0 0 0];end ;
        if (isempty(VQuad5)),VQuad5=[0 0 0];end ;
        if (isempty(VQuad6)),VQuad6=[0 0 0];end ;
        if (isempty(VQuad7)),VQuad7=[0 0 0];end ;
        %% Statistics Octant Feature Extraction
        Quad7Num = size(VQuad7,1);Quad6Num = size(VQuad6,1);
        Quad5Num = size(VQuad5,1);Quad4Num = size(VQuad4,1);
```

```
        Quad3Num = size(VQuad3,1);Quad2Num = size(VQuad2,1);
        Quad1Num = size(VQuad1,1);Quad0Num = size(VQuad0,1);

        Quad7Norm= sqrt(sum(VQuad7.^2,2));Quad6Norm=
sqrt(sum(VQuad6.^2,2));
        Quad5Norm= sqrt(sum(VQuad5.^2,2));Quad4Norm=
sqrt(sum(VQuad4.^2,2));
        Quad3Norm= sqrt(sum(VQuad3.^2,2));Quad2Norm=
sqrt(sum(VQuad2.^2,2));
        Quad1Norm= sqrt(sum(VQuad1.^2,2));Quad0Norm=
sqrt(sum(VQuad0.^2,2));


        Quad7MaxN = max(Quad7Norm);Quad7AvgN =
mean(Quad7Norm);Quad7VarN = var(Quad7Norm);
        Quad6MaxN = max(Quad6Norm);Quad6AvgN =
mean(Quad6Norm);Quad6VarN = var(Quad6Norm);
        Quad5MaxN = max(Quad5Norm);Quad5AvgN =
mean(Quad5Norm);Quad5VarN = var(Quad5Norm);
        Quad4MaxN = max(Quad4Norm);Quad4AvgN =
mean(Quad4Norm);Quad4VarN = var(Quad4Norm);
        Quad3MaxN = max(Quad3Norm);Quad3AvgN =
mean(Quad3Norm);Quad3VarN = var(Quad3Norm);
        Quad2MaxN = max(Quad2Norm);Quad2AvgN =
mean(Quad2Norm);Quad2VarN = var(Quad2Norm);
        Quad1MaxN = max(Quad1Norm);Quad1AvgN =
mean(Quad1Norm);Quad1VarN = var(Quad1Norm);
        Quad0MaxN = max(Quad0Norm);Quad0AvgN =
mean(Quad0Norm);Quad0VarN = var(Quad0Norm);

        quad_stat=[Quad0Num/5000 Quad0MaxN Quad0AvgN Quad0VarN ...  %
Statistics octant feature matrix
            Quad1Num/5000 Quad1MaxN Quad1AvgN Quad1VarN ...
            Quad2Num/5000 Quad2MaxN Quad2AvgN Quad2VarN ...
            Quad3Num/5000 Quad3MaxN Quad3AvgN Quad3VarN ...
            Quad4Num/5000 Quad4MaxN Quad4AvgN Quad4VarN ...
            Quad5Num/5000 Quad5MaxN Quad5AvgN Quad5VarN ...
            Quad6Num/5000 Quad6MaxN Quad6AvgN Quad6VarN ...
            Quad7Num/5000 Quad7MaxN Quad7AvgN Quad7VarN];

%% Polar Coordinate Octant Features

           [MaxOct7,Ind7] = max(Quad7Norm);[MaxOct6,Ind6] =
max(Quad6Norm);
           [MaxOct5,Ind5] = max(Quad5Norm);[MaxOct4,Ind4] =
max(Quad4Norm);
           [MaxOct3,Ind3] = max(Quad3Norm);[MaxOct2,Ind2] =
max(Quad2Norm);
           [MaxOct1,Ind1] = max(Quad1Norm);[MaxOct0,Ind0] =
max(Quad0Norm);
```

```
            [THETA7_MAX,PHI7_MAX,R7_MAX]=
cart2sph(VQuad7(Ind7,1),VQuad7(Ind7,2),VQuad7(Ind7,3));
            [THETA6_MAX,PHI6_MAX,R6_MAX]=
cart2sph(VQuad6(Ind6,1),VQuad6(Ind6,2),VQuad6(Ind6,3));
            [THETA5_MAX,PHI5_MAX,R5_MAX]=
cart2sph(VQuad5(Ind5,1),VQuad5(Ind5,2),VQuad5(Ind5,3));
            [THETA4_MAX,PHI4_MAX,R4_MAX]=
cart2sph(VQuad4(Ind4,1),VQuad4(Ind4,2),VQuad4(Ind4,3));
            [THETA3_MAX,PHI3_MAX,R3_MAX]=
cart2sph(VQuad3(Ind3,1),VQuad3(Ind3,2),VQuad3(Ind3,3));
            [THETA2_MAX,PHI2_MAX,R2_MAX]=
cart2sph(VQuad2(Ind2,1),VQuad2(Ind2,2),VQuad2(Ind2,3));
            [THETA1_MAX,PHI1_MAX,R1_MAX]=
cart2sph(VQuad1(Ind1,1),VQuad1(Ind1,2),VQuad1(Ind1,3));
            [THETA0_MAX,PHI0_MAX,R0_MAX]=
cart2sph(VQuad0(Ind0,1),VQuad0(Ind0,2),VQuad0(Ind0,3));
          pol_stat= [THETA0_MAX PHI0_MAX R0_MAX THETA1_MAX PHI1_MAX
R1_MAX...
              THETA2_MAX PHI2_MAX R2_MAX THETA3_MAX PHI3_MAX R3_MAX...
              THETA4_MAX PHI4_MAX R4_MAX THETA5_MAX PHI5_MAX R5_MAX...
              THETA6_MAX PHI6_MAX R6_MAX THETA7_MAX PHI7_MAX R7_MAX];%
Polar

%% Octant distribution

          x=1:1:length(VCG(:,1));
          y=1:1:length(VCG(:,2));
          z=1:1:length(VCG(:,3));
          [theta,phi,radius]= cart2sph(VCG(:,1),VCG(:,2),VCG(:,3));
%to sphere coordinator
          radius_ext =radius*5;
          deadzone = 0.05;
          median_order = round(0.05*avrRR*fs); % window for median filter
          if rem(median_order,2)==0
              median_order = median_order +1;  %create an odd median
window
          end
          c1=((VCG(:,1)<0) & (VCG(:,2)<0) & (VCG(:,3)<0) & (radius >
deadzone*rr_series')); %Form the color coding
          c2=((VCG(:,1)<0) & (VCG(:,2)<0) & (VCG(:,3)>0) & (radius >
deadzone*rr_series'));
          c3=((VCG(:,1)<0) & (VCG(:,2)>0) & (VCG(:,3)<0) & (radius >
deadzone*rr_series'));
          c4=((VCG(:,1)<0) & (VCG(:,2)>0) & (VCG(:,3)>0) & (radius >
deadzone*rr_series'));
          c5=((VCG(:,1)>0) & (VCG(:,2)<0) & (VCG(:,3)<0) & (radius >
deadzone*rr_series'));
          c6=((VCG(:,1)>0) & (VCG(:,2)<0) & (VCG(:,3)>0) & (radius >
deadzone*rr_series'));
          c7=((VCG(:,1)>0) & (VCG(:,2)>0) & (VCG(:,3)<0) & (radius >
deadzone*rr_series'));
          c8=((VCG(:,1)>0) & (VCG(:,2)>0) & (VCG(:,3)>0) & (radius >
deadzone*rr_series'));
```

```
        c=1*c1+2*c2+3*c3+4*c4+5*c5+6*c6+7*c7+8*c8;
        c = medfilt1(c,median_order); % octant transition noise
        mindiffC= min(diff(c));
        maxdiffC= max(diff(c));
        meandiffC=mean(diff(c));
        meanC=mean(c);
        oct_dist=[mindiffC maxdiffC meandiffC meanC]; %Octant
distribution
        %% Fomulate the transition series
        nonzero_ind = find(diff(c));
        tran_ind = [ c(nonzero_ind);c(nonzero_ind(end)+1)];

%% Transition rate matrix formulation

        c1dec=c1;
        c2dec=2*c2;
        c3dec=3*c3;
        c4dec=4*c4;
        c5dec=5*c5;
        c6dec=6*c6;
        c7dec=7*c7;
        c8dec=8*c8;
        c12=diff(c2dec)-diff(c1dec); %transition 1-2 => c12=3
        c13=diff(c3dec)-diff(c1dec); %transition 1-3 => c13=4
        c14=diff(c4dec)-diff(c1dec);
        c15=diff(c5dec)-diff(c1dec);
        c16=diff(c6dec)-diff(c1dec);
        c17=diff(c7dec)-diff(c1dec);
        c18=diff(c8dec)-diff(c1dec);
        c23=diff(c3dec)-diff(c2dec);
        c24=diff(c4dec)-diff(c2dec);
        c25=diff(c5dec)-diff(c2dec);
        c26=diff(c6dec)-diff(c2dec);
        c27=diff(c7dec)-diff(c2dec);
        c28=diff(c8dec)-diff(c2dec);
        c34=diff(c4dec)-diff(c3dec);
        c35=diff(c5dec)-diff(c3dec);
        c36=diff(c6dec)-diff(c3dec);
        c37=diff(c7dec)-diff(c3dec);
        c38=diff(c8dec)-diff(c3dec);
        c45=diff(c5dec)-diff(c4dec);
        c46=diff(c6dec)-diff(c4dec);
        c47=diff(c7dec)-diff(c4dec);
        c48=diff(c8dec)-diff(c4dec);
        c56=diff(c6dec)-diff(c5dec);
        c57=diff(c7dec)-diff(c5dec);
        c58=diff(c8dec)-diff(c5dec);
        c67=diff(c7dec)-diff(c6dec);
        c68=diff(c8dec)-diff(c6dec);
        c78=diff(c8dec)-diff(c7dec);
        %Changing rate
        t_n1=sum([(c12==-3);(c13==-4);(c14==-5);(c15 ==-6);...
```

```
            (c16 ==-7);(c17==-8);(c18==-9)]); %number of transition to
1
        t_n2=sum([(c12==3);(c23==-5);(c24==-6);(c25 ==-7);...
            (c26 ==-8);(c27==-9);(c28==-10)]);
        t_n3=sum([(c13==4);(c23==5);(c34==-7);(c35 ==-8);...
            (c36 ==-9);(c37==-10);(c38==-11)]);
        t_n4=sum([(c14==5);(c24==6);(c34==7);(c45 ==-9);...
            (c46 ==-10);(c47==-11);(c48==-12)]);
        t_n5=sum([(c15==6);(c25==7);(c35==8);(c45 ==9);...
            (c56 ==-11);(c57==-12);(c58==-13)]);
        t_n6=sum([(c16==7);(c26==8);(c36==9);(c46 ==10);...
            (c56 == 11);(c67==-13);(c68==-14)]);
        t_n7=sum([(c17==8);(c27==9);(c37==10);(c47 ==11);...
            (c57 == 12);(c67==13);(c78==-15)]);
        t_n8=sum([(c18==9);(c28==10);(c38==11);(c48 ==12);...
            (c58 == 13);(c68==14);(c78==15)]);
        t_1n=sum([(c12==3);(c13==4);(c14==5);(c15==6);...
            (c16==7);(c17==8);c(18==9)]);            %number of
transitions from 1, if from 1-2->c12=3
        t_2n=sum([(c12==-3);(c23==5);(c24==6);(c25==7);...
            (c26==8);(c27==9);(c28==10)]);
        t_3n=sum([(c13==-4);(c23==-5);(c34==7);(c35==8);...
            (c36==6);(c37==10);(c38==11)]);
        t_4n=sum([(c14==-5);(c24==-6);(c34==-7);(c45==9);...
            (c46==10);(c47==11);(c48==12)]);
        t_5n=sum([(c15==-6);(c25==-7);(c35==-8);(c45 ==-9);...
            (c56 ==11);(c57==12);(c58==13)]);
        t_6n=sum([(c16==-7);(c26==-8);(c36==-9);(c46 ==-10);...
            (c56 == -11);(c67==13);(c68==14)]);
        t_7n=sum([(c17==-8);(c27==-9);(c37==-10);(c47 ==-11);...
            (c57 ==-12);(c67==-13);(c78==15)]);
        t_8n=sum([(c18==-9);(c28==-10);(c38==-11);(c48 ==-12);...
            (c58 ==-13);(c68==-14);(c78==-15)]);
        tran_rate= [t_n1 t_n2 t_n3 t_n4 t_n5 t_n6 t_n7 t_n8...
            t_1n t_2n t_3n t_4n t_5n t_6n t_7n t_8n];

 %% Adjacent matrix with 64 elements

        %To 1 and from 1
        a12=sum(diff(c2dec)-diff(c1dec)==3); %transition 1-2 => c12=3
        a13=sum(diff(c3dec)-diff(c1dec)==4); %transition 1-3 => c13=4
        a14=sum(diff(c4dec)-diff(c1dec)==5);
        a15=sum(diff(c5dec)-diff(c1dec)==6);
        a16=sum(diff(c6dec)-diff(c1dec)==7);
        a17=sum(diff(c7dec)-diff(c1dec)==8);
        a18=sum(diff(c8dec)-diff(c1dec)==9);
        a21=sum(diff(c2dec)-diff(c1dec)==-3); %transition 1-2 => c12=3
        a31=sum(diff(c3dec)-diff(c1dec)==-4); %transition 1-3 => c13=4
        a41=sum(diff(c4dec)-diff(c1dec)==-5);
        a51=sum(diff(c5dec)-diff(c1dec)==-6);
        a61=sum(diff(c6dec)-diff(c1dec)==-7);
        a71=sum(diff(c7dec)-diff(c1dec)==-8);
        a81=sum(diff(c8dec)-diff(c1dec)==-9);
```

```
        %To 2 and from 2
        a23=sum(diff(c3dec)-diff(c2dec)==5);
        a24=sum(diff(c4dec)-diff(c2dec)==6);
        a25=sum(diff(c5dec)-diff(c2dec)==7);
        a26=sum(diff(c6dec)-diff(c2dec)==8);
        a27=sum(diff(c7dec)-diff(c2dec)==9);
        a28=sum(diff(c8dec)-diff(c2dec)==10);
        a32=sum(diff(c3dec)-diff(c2dec)==-5);
        a42=sum(diff(c4dec)-diff(c2dec)==-6);
        a52=sum(diff(c5dec)-diff(c2dec)==-7);
        a62=sum(diff(c6dec)-diff(c2dec)==-8);
        a72=sum(diff(c7dec)-diff(c2dec)==-9);
        a82=sum(diff(c8dec)-diff(c2dec)==-10);
        %To 3 and from 3
        a34=sum(diff(c4dec)-diff(c3dec)==7);
        a35=sum(diff(c5dec)-diff(c3dec)==8);
        a36=sum(diff(c6dec)-diff(c3dec)==9);
        a37=sum(diff(c7dec)-diff(c3dec)==10);
        a38=sum(diff(c8dec)-diff(c3dec)==11);
        a43=sum(diff(c4dec)-diff(c3dec)==-7);
        a53=sum(diff(c5dec)-diff(c3dec)==-8);
        a63=sum(diff(c6dec)-diff(c3dec)==-9);
        a73=sum(diff(c7dec)-diff(c3dec)==-10);
        a83=sum(diff(c8dec)-diff(c3dec)==-11);
        %To 4 and from 4
        a45=sum(diff(c5dec)-diff(c4dec)==9);
        a46=sum(diff(c6dec)-diff(c4dec)==10);
        a47=sum(diff(c7dec)-diff(c4dec)==11);
        a48=sum(diff(c8dec)-diff(c4dec)==12);
        a54=sum(diff(c5dec)-diff(c4dec)==-9);
        a64=sum(diff(c6dec)-diff(c4dec)==-10);
        a74=sum(diff(c7dec)-diff(c4dec)==-11);
        a84=sum(diff(c8dec)-diff(c4dec)==-12);
        %To 5 and from 5
        a56=sum(diff(c6dec)-diff(c5dec)==11);
        a57=sum(diff(c7dec)-diff(c5dec)==12);
        a58=sum(diff(c8dec)-diff(c5dec)==13);
        a65=sum(diff(c6dec)-diff(c5dec)==-11);
        a75=sum(diff(c7dec)-diff(c5dec)==-12);
        a85=sum(diff(c8dec)-diff(c5dec)==-13);
        %To 6 and from 6
        a67=sum(diff(c7dec)-diff(c6dec)==13);
        a68=sum(diff(c8dec)-diff(c6dec)==14);
        a76=sum(diff(c7dec)-diff(c6dec)==-13);
        a86=sum(diff(c8dec)-diff(c6dec)==-14);
        %To 7 and from 7
        a78=sum(diff(c8dec)-diff(c7dec)==15);
        a87=sum(diff(c8dec)-diff(c7dec)==-15);
        a11=0;a22=0;a33=0;a44=0;a55=0;a66=0;a77=0;a88=0;
        adjMat=[a11 a12 a13 a14 a15 a16 a17 a18;...  %formulate the
transition matrix in adjacency matrix
            a21 a22 a23 a24 a25 a26 a27 a28;...
            a31 a32 a33 a34 a35 a36 a37 a38;...
            a41 a42 a43 a44 a45 a46 a47 a48;...
```

```
            a51 a52 a53 a54 a55 a56 a57 a58;...
            a61 a62 a63 a64 a65 a66 a67 a68;...
            a71 a72 a73 a74 a75 a76 a77 a78;...
            a81 a82 a83 a84 a85 a86 a87 a88];
        tempMat = adjMat;
         %Extract feature from the transition matrix
         % Degree and Assortativity
         [id,od,deg] = degrees_dir(tempMat); % node degree
         [is,os,str] = strengths_dir(tempMat); % node strength
         [J,J_od,J_id,J_bl] = jdegree(tempMat); % join degree
J_od,J_id,J_bl number of vertices with od>id,id>od,id=od.
         r_ass = assortativity_bin(tempMat,1);
         deg_assort =[id,od,deg,is,os,str,J_od,J_id,J_bl,r_ass];
         % Density,Clustering, and Modularity
         [kden,n_den,k_den] = density_dir(tempMat); % Densitivity
         Cluster=clustering_coef_wd(tempMat); %Clustering coeff
         Cluster=Cluster';% create a row vector *
         transi=transitivity_wd(tempMat);
         [ci_mod q_mod]=modularity_dir(tempMat); %(use  modularity in
the old version )
         den_clust_mod = [kden,n_den,k_den,Cluster,transi,q_mod];
         %Distances and Cycles
         dis_weight=distance_wei(tempMat); %distant weight matrix **
just used to cal mean_dis_weight
         [lambda_net,efficiency_net,ecc,radius_net,diameter_net] =
charpath(dis_weight);% path characteristics
         ecc =ecc';

dis_cyc=[lambda_net,efficiency_net,ecc,radius_net,diameter_net];
         %Centrality
         node_bet=betweenness_wei(tempMat);
         node_bet=node_bet';
         centr = node_bet;
         %Sum up the network features
         network_mea =[deg_assort,den_clust_mod,dis_cyc,centr];
        end
end
```

A.4 CART classification

```
%%Classification model

dbstop if error; %Stop at the error
Path='ECG_Features_MILocalization.xlsx';
idx = xlsread(Path,'cart','B2:B363'); % with the network measure
features
[num1,diag1] = xlsread(Path,'cart','E2:E363');
% [num2,loc] = xlsread('VCGcart.xls','cart','G2:G363');
[num3,name] = xlsread(Path,'cart','F1:CC1');
B = xlsread(Path,'cart','F2:Y363');
n1 = 0; n2=0;n3=0;n4=0;n5=0;n6=0;
for i = 1:length(diag1)
```

```
    if (strcmp(diag1(i),'Infero') ||strcmp(diag1(i),'Infero Lateral')
||...
            strcmp(diag1(i),'Infero Posterior')||
strcmp(diag1(i),'Infero Postero Lateral'))
        n1 =n1+1;
        Inf(n1,:)=B(i,:);
    end
  if (strcmp(diag1(i),'Anterior')||strcmp(diag1(i),'Anterio
Septal')||strcmp(diag1(i),'Antero Lateral'))
        n2=n2+1;
        Ant(n2,:)=B(i,:);
    end

    if strcmp(diag1(i),'HC')
        n3=n3+1;
        HCtl(n3,:)=B(i,:);
    end
end

%% Features selection

%K-sample Anderson Darling test
for i=1:1:min(size(B))
    i
    ADTest = [];
    ADData = [Inf(1:n1,i);Ant(1:n2,i);HCtl(1:n3,i)];% first column of
data
    ADGrp  = [1*ones(n1,1);2*ones(n2,1);3*ones(n3,1)];% second column
of group sample
    ADTest =[ADData ADGrp];
    if i ==64
        ADPn(64) = ADPn(63);
        ADP(64) = ADP(63);
        ADhyp(64)=ADhyp(63);
    else if i==134
            ADPn(134) = ADPn(133);
            ADP(134) = ADP(133);
            ADhyp(134)=ADhyp(133);
        else
            [ADPn(i),ADP(i),ADhyp(i)] = AnDarksamtest(ADTest,0.01);
        end
    end
end
for i=1:1:min(size(B))
    KSData={Inf(1:n1,i);Ant(1:n2,i);HCtl(1:n3,i)};
    p=[];h=[];k=[];
    count=0;
    for step1 =1:1:2
        for step2 =step1+1:1:3
            count = count+1;
            [h(count),p(count),k(count)] =
kstest2(KSData{step1,1},KSData{step2,1});
        end
```

```
    end
    KS(i) = sum(p);
    KS_R(i)= sum(h);
    KSSTAT(i)=sum(k);
end
[b2,imp2] = sort(KS,'ascend');  %Selectetd by KS test

Feature
={'QI';'QII';'QIII';'QAVL';'QAVF';'QV1';'QV2';'QV3';'QV4';'QV5';...

'RI';'RII';'RIII';'RAVL';'RAVF';'RV1';'RV2';'RV3';'RV4';'RV5'};
A=[ADP' ADPn' ADhyp' KS' KS_R' KSSTAT'];
imp1=imp2(1:10);
Feature1= Feature(imp2)';
Feature = Feature(imp1);
Inf = Inf(:,imp1);                % Feature matrix of inferior group
InfAnno(1:n1,1) = cellstr('Infero'); %Annotation of inferior group
Ant = Ant(:,imp1);
AntAnno(1:n2,:) = cellstr('Anterior');
HCtl = HCtl(:,imp1);
HCtlAnno(1:n3,:)= cellstr('HC');
permtx = [10/9 1.25 1.5 2 3 4 5 10]; %
(10%,20%,33.33%,50%,66.67%,75%,80%,90%)
 permtx = 10; %90% for training
C=zeros(3);
for ipercent = 1:8
    percent = permtx(ipercent);
    for i = 1:20
        iindex = randperm(n1); % K-fold cross validation
        aindex = randperm(n2);
        hcindex = randperm(n3);
        %Training data
        trn_i = Inf(iindex(floor(n1/percent)+1:n1),:); %training data
percent =10/9-> 10% for training
        trn_i_anno = InfAnno(iindex(floor(n1/percent)+1:n1),1);%
training annotation group
        trn_a = Ant(aindex(floor(n2/percent)+1:n2),:);
        trn_a_anno = AntAnno(aindex(floor(n2/percent)+1:n2),:);
        trn_hc =HCtl(hcindex(floor(n3/percent)+1:n3),:);
        trn_hc_anno =HCtlAnno(hcindex(floor(n3/percent)+1:n3),:);
        trn_data = [trn_i;trn_a;trn_hc]; % Formulate the training data
set
        trn_anno = [trn_i_anno;trn_a_anno;trn_hc_anno]; %Training
annotation
        %Testing data
        test_i = Inf(iindex(1:floor(n1/percent)),:); %Testing data
percent =10/9-> 90% for testing
        test_i_anno = InfAnno(iindex(1:floor(n1/percent)),1);% training
annotation group
        test_a = Ant(aindex(1:floor(n2/percent)),:);
        test_a_anno = AntAnno(aindex(1:floor(n2/percent)),:);
        test_hc =HCtl(hcindex(1:floor(n3/percent)),:);
        test_hc_anno = HCtlAnno(hcindex(1:floor(n3/percent)),:);
```

```
        test_data = [test_i;test_a;test_hc]; % Formulate the Testing
data set
        test_anno = [test_i_anno;test_a_anno;test_hc_anno]; %Testing
annotation
        whole_data=[Inf;Ant;HCtl];
        whole_anno=[InfAnno;AntAnno;HCtlAnno];
        numobs = size(test_anno,1);
        tree = treefit(trn_data,trn_anno);  %Fit the tree
        [c,s,n,best] = treetest(tree,'cross',trn_data,trn_anno);
        tmin = treeprune(tree,'level',best);
            treedisp(tmin,'names',Feature);
        [dtnum,dtnode,dtclass] = treeval(tmin, test_data); %
Classification with the testing data
        bad = ~strcmp(dtclass,test_anno);
        c1= length(test_i_anno);c2= length(test_a_anno);c3=
length(test_hc_anno);
        abad(i) = sum(bad(c1+1:c1+c2));
        hcbad(i) = sum(bad(c1+c2+1:c1+c2+c3));
        acc_i(i) = 100-100*ibad(i)/(c1);
        acc_a(i) = 100-100*abad(i)/(c2);
        acc_hc(i) = 100-100*hcbad(i)/(c3);
        acc_total(i) = 100-100*sum(bad)/numobs;
        C = C+confusionmat(test_anno,dtclass);

        fprintf('.....Percent: %d ..rand: %d ..Inf: %d ..Ant:%d ..HC:%d
..\n',...
            ipercent,i,acc_i(i),acc_a(i),acc_hc(i));
    end
    Cmean = C./i;
    restat = [1-1/percent  mean(acc_i) std(acc_i)...
         mean(acc_a) std(acc_a)...
        mean(acc_hc) std(acc_hc) mean(acc_total) std(acc_total)];
```

A.5 Dirichlet process based Gaussian process mixture (DPMG) prediction

```
HistPSD=(1:10);
HistLVM=(1:10);
ECGSignal=data;
fs=100;
PeakIdx = nqrsdetect(ECGSignal,fs); %Identify QRS using filter bank
method

%% Power spectrum density (PSD) feature from RR interval

    RR = diff(PeakIdx)/fs;                          %RR interval
    RR=[0;RR];     %Balance with PeakIDx
    RR_Peak=[PeakIdx RR];
    ModifiedRR = clean_hrv4(RR_Peak); %default threshold is 20%
difference ;
    PeakIdx1= ModifiedRR(:,1);
    RR1= ModifiedRR(:,2);
    RR_down = interp1(PeakIdx1,RR1,min(PeakIdx1):fs/2:max(PeakIdx1));
```

```
    FilterRR = Filter1(RR_down);                              %Get 0.02 to
0.1 Hz
    [S,F,T,P]= spectrogram(FilterRR,length(FilterRR),0,120,2);
%Spectrogram
     CurrPSD = RMS1(P);   %PSD of the current minute

%% Longest vertical line (LVM) from RR interval using RQA

Extract the longest vertical line property
RR = spline(PeakIdx1,RR1,1:fs:PeakIdx1(end,1));
dim=7;
Delay=5;
[RRy] = PhaseSpace(RR,dim,Delay);
Amplitude=[];
for step1 =1:1:max(size((RRy)))
    Amplitude(step1)=norm(RRy(step1,:));
end
Thresh = 0.1*(max(Amplitude)- min(Amplitude));
RQA =crqa(RR,7,5,Thresh,60*10,60*1,3,3,1);
CurrVLM =RQA(8);% Max of Vertical Lenghth of the current minute

%% Prediction of PSD feature

ParameterEst();
HistPSD = (HistPSD-Par1)/(Par2-Par1);
HistPSD = (HistPSD -Par3)/Par4;
XnPSD =PhaseSpace(HistPSD,dd1,tau1);
XnPSD=XnPSD(end,:)'; %Get the last row for the phase space information
of the current point

for k=1:1:length(pp1)  % k clusters from the training process
    pp1(k)=mvnpdf(XnPSD,mu1{k},sigma1{k});    % Likelihood of the Xn
    hp=exp((logtheta1{k})');
    [fstar S2] = gpra(logtheta1{k}, covfunc, X1{k}, y1{k}, XnPSD');
    S2 = S2 - exp(2*logtheta1{k}(dim1(k)));   % 95% confidence interval
    fm1(k) = fstar;
    fvar1(k) = S2;
end
pb1 = pp1/sum(pp1);
PSD_pred = sum(pb1.*fm1);

%% Prediction of LVM feature

Data2LVM = Data2LVM';
datLVM = Data2LVM(4:end,1:end-1);
datLVM0 = Data2LVM(4:end,1:end);
dimenLVM = 5;  %dimension of one cluster
dpm2 = dpm_gp(datLVM,dimenLVM,aa, s0, ss, numiter);
pp2 = zeros(dpm2.KK,1);
fm2 = zeros(dpm2.KK,1);
fvar2 = zeros(dpm2.KK,1);
% Training %
```

```
for j=1:dpm2.KK
    [mu2{j},sigma2{j}] = map(dpm2.qq{j});  %% Estimated mean and
variance using MAP estimation
    j
    %     cluster{j} = find(dpm.zz==j);
    id = find(dpm2.zz==j);
    X2{j} = datLVM(:,id)';
    y2{j} = datLVM0(dd1,id+1)';    % output of the prediction
    covfunc = {'covSum',{'covConst','covSEard','covNoise'}};
    [n,D]=size(X2{j});
    dim2(j)=eval(feval(covfunc{:}));
    logtheta02 =zeros(dim2(j),1); % initial value
    logtheta02(1:dim2(j)-2)=log(1);    % the characteristic length must
be large according to the large variance
    logtheta02(1)=log(1);
    logtheta02(dim2(j)-1)=log(1);
    logtheta02(dim2(j))=log(0.01);
    [logtheta12{j}, fvals2{j}, iter2{j}] = minimize(logtheta02, 'gpra',
-100, covfunc, X2{j}, y2{j}); % optimize the hyperparameter
end
XnLVM =PhaseSpace(HistLVM,dd1,tau1);
XnLVM=XnLVM(end,:)'; %Get the last row for the phase space information
of the current point

for k=1:1:length(pp2)  % k clusters from the training process
    pp2(k)=mvnpdf(XnLVM,mu2{k},sigma2{k});    % Likelihood of the Xn
%     disp('  exp(loghyper) =');
%     disp(exp(logtheta{k})');
    hp=exp((logtheta12{k})');
    [fstar2 S22] = gpra(logtheta12{k}, covfunc, X2{k}, y2{k}, XnLVM');
    S22 = S22 - exp(2*logtheta12{k}(dim2(k)));   % 95% confidence
interval
    fm2(k) = fstar2;
    fvar2(k) = S22;
end
pb2 = pp2/sum(pp2);
LVM_pred = sum(pb2.*fm2);

%% Classification

PredValues =[PSD_pred*Par4+Par3];
Data2PSD =Data2PSD';
Apnea = Data2PSD(Data2PSD(:,1)==1,3);
NonApnea= Data2PSD(Data2PSD(:,1)==0,3);
options = optimset('maxiter', 1000, 'largescale','off');
sigma = exp(-0.7533);
constr = exp(-0.75);
svmStruct =
svmtrain([Apnea;NonApnea],[ones(size(Apnea,1),1);zeros(size(NonApnea,1)
,1)],'Kernel_Function','rbf',...
    'rbf_sigma',sigma,'boxconstraint',constr);
ApnStatus = svmclassify(svmStruct,PredValues);
```

```
ApnStatus =
classify(PredValues,[Apnea;NonApnea],[ones(size(Apnea,1),1);zeros(size(
NonApnea,1),1)],'diagquadratic');

function ParameterEst()
aa = 1;
s0 = 1;
ss = 0.5;
numiter = 100;
tau1 =2; %  time delay for both PSD and LVM
dd1 = 5; % dimension of state space
Par1 = 3.2302e-004; %Min value of PSD for normalization
Par2 = 1.1949; % Max value of PSD for morlization
load 'L:\Research\Sleep Apnea Prediction Garment\Predition
Apnea\Changqing Code\DPGM\10ApnPatients.mat'; %Select the
dbstop if error;
for i= 1:1:length(Data)
%     Temp = [];
    TempPSD = Data{i}(:,3);
    TempPSD = (TempPSD-mean(TempPSD))/std(TempPSD); %Normalize PSD
    AvgPSD(i)= mean(TempPSD);
    StdPSD(i)= std(TempPSD);
    TempLVM= Data{i}(:,2);
    TempLVM = (TempLVM-mean(TempLVM))/std(TempLVM);
    NPSD=PhaseSpace(TempPSD,dd1,tau1);
    LVM=PhaseSpace(TempLVM,dd1,tau1);
    Data1PSD{i}=[Data{i}(1+2*(dd1-1):end,1:3) NPSD]; % Col1: Apnea
Status, Col2: LVM, Col3: NPSD, Col4-8: Phase space information of PSD
    Data1LVM{i}= [Data{i}(1+2*(dd1-1):end,1:3) LVM];  % Col1: Apnea
Status, Col2: LVM, Col3: NPSD, Col4-8: Phase space information of LVM
end
Par3 = mean(AvgPSD);
Par4 = mean(StdPSD);
index =[1:1:length(Data)];
TestId = [];  % Testing data
index(TestId)=[]; % Keep it for testing
Data2PSD=[];
Data2LVM=[];
for i=index
    Data2PSD = [Data2PSD;Data1PSD{i}]; %Whole data set
    Data2LVM = [Data2LVM;Data1LVM{i}];
end
Data2PSD = Data2PSD';
datPSD = Data2PSD(4:end,1:end-1);
datPSD0 = Data2PSD(4:end,1:end);
dimenPSD = 5;  %dimension of one cluster
dpm1 = dpm_gp(datPSD,dimenPSD,aa, s0, ss, numiter);
pp1 = zeros(dpm1.KK,1);
fm1 = zeros(dpm1.KK,1);
fvar1 = zeros(dpm1.KK,1);
% Training %
for j=1:dpm1.KK
```

```
    [mu1{j},sigma1{j}] = map(dpm1.qq{j});  %% Estimated mean and
variance using MAP estimation
    id = find(dpm1.zz==j);
    X1{j} = datPSD(:,id)';
    y1{j} = datPSD0(dd1,id+1)';    % output of the prediction
    covfunc = {'covSum',{'covConst','covSEard','covNoise'}};
    [n,D]=size(X1{j});
    dim1(j)=eval(feval(covfunc{:}));
    logtheta0 =zeros(dim1(j),1); % initial value
    logtheta0(1:dim1(j)-2)=log(1);    % the characteristic length must
be large according to the large variance
    logtheta0(1)=log(1);
    logtheta0(dim1(j)-1)=log(1);
    logtheta0(dim1(j))=log(0.01);
    [logtheta1{j}, fvals1{j}, iter1{j}] = minimize(logtheta0, 'gpra', -
100, covfunc, X1{j}, y1{j}); % optimize the hyperparameter
end
save all;
end


function PredictDPMG(data)
avg = mean(data);
stdv = std(data);
data = (data-avg)/stdv;
dbstop if error;
aa = 0.6;
s0 = 1;
ss = 0.5;
numiter = 5;
tau =2; %  time delay
dd = 5; % dimension of state space
SP = phasespaceh(data,dd,tau);
dat0 = SP';
dat = dat0(:,1:500);
Xn = dat0(:,501);
f = dat0(5,502);
dim = size(dat,1);
dpm = dpm_gp(dat,dim,aa, s0, ss, numiter);
pp = zeros(dpm.KK,1);
% cluster = cell(dpm.KK,1);
fm = zeros(dpm.KK,1);
fvar = zeros(dpm.KK,1);
for j=1:dpm.KK
    [mu,sigma] = map(dpm.qq{j});  %% Estimated mean and variance using
MAP estimation
    pp(j)=mvnpdf(Xn,mu,sigma);
    %     cluster{j} = find(dpm.zz==j);
    id = find(dpm.zz==j);
    X = dat(:,id)';
    y = dat0(dd,id+1)';
    covfunc = {'covSum',{'covConst','covSEard','covNoise'}};
    [n,D]=size(X);
     dim=eval(feval(covfunc{:}));
```

```
    logtheta0 =zeros(dim,1); % initial value
    logtheta0(1:dim-2)=log(1);    % the characteristic length must be
large according to the large variance
    logtheta0(1)=log(1);
    logtheta0(dim-1)=log(1);
    logtheta0(dim)=log(0.01);
    [logtheta, fvals, iter] = minimize(logtheta0, 'gpra', -100,
covfunc, X, y); % optimize the hyperparameter
    disp('  exp(loghyper) =')
    disp(exp(logtheta)')
    hp=exp((logtheta)');
    [fstar S2] = gpra(logtheta, covfunc, X, y, Xn');
    S2 = S2 - exp(2*logtheta(dim));   % 95% confidence interval
    fm(j) = fstar;
    fvar(j) = S2;
end
pb = pp/sum(pp);
f_pred = sum(pb.*fm);
disp([f f_pred])
end
```

A.6 Color coded state space network representation

```
%% Prepare the data

load L:\Research\Hoang-Prognosis\MatlabCodes\10ApnPatients.mat; %Sleep
apnea annotation, NPSD, and LVM
load('L:\Research\Hoang-Prognosis\MatlabCodes\TimeDelay.mat') %Time
delay file, identified by using MI
 for patientId=2
    objId = patientId;
    dirpath= strcat('L:\Research\Hoang-
Prognosis\MatlabCodes\DataGood_Feature1\','ID',num2str(objId),'\');
    data = Data{objId}(:,3); % Power spectrum density
    Slp_ann = Data{objId}(:,1);
    avg = mean(data);
    stdv = std(data);
    SleepAnno = Data{objId}(:,1);
    data = (data-avg)/stdv;
    LVM1 = Data{objId}(:,2); %Longest vertical length
    avg1 = mean(LVM1);
    stdv1 = std(LVM1);
    LVM = (LVM1-avg1)/stdv;

%% Multivariate state space resconstruction

    TauPSD = TimeDelay(patientId,1);
    TauLVM = TimeDelay(patientId,2);

    % tunable parameters, aa, s0, ss

    aa = 0.6;
    s0 = 1;
```

```
    ss = 0.5;
    numiter = 10;
    radius = 0.1; %radius of the neighbors 5% around the new point
    DimPSD = 10;
    DimLVM = 10;
    N= length(data);
    minimumx = min(LVM1);
    maximumx = max(LVM1);
    intervalx = maximumx-minimumx;
    LVM1=(LVM1-minimumx)/intervalx;
    avg1 = mean(LVM1);
    stdv1 = std(LVM1);
    LVM = (LVM1-avg1)/stdv;
    NX1 = N-TauPSD*(DimPSD-1);
    NX2 = N-TauLVM*(DimLVM-1);
    NX = min(NX1,NX2);
    phase1 = PhaseSpace(data,DimPSD,TauPSD);
    phase2 = PhaseSpace(LVM,DimLVM,TauLVM);
    SP = [phase1(1:NX,:),phase2(1:NX,:)];
    d = Lyapunov_exp(SP,5,500);
    diaPhase=[];
    dat0 = SP';
    train_point = 300; % number of minimum training points
    test_point1 = train_point-1 + ceil((NX-train_point) * rand(NX-
train_point,1));  % Select extra random testing point from 300th point
    apn_point= test_point1(Slp_ann(test_point1)==1);
    non_point= test_point1(Slp_ann(test_point1)==0);
    test_point = non_point; % for non points (selelec either this one
or the apnea points)
    if (~isempty(test_point))
    t_pred_allr=[];
    t_std_pred_allr=[];
    ChangeInt1_allr=[];
    t_pred1_allr=[];
    t_std_pred1_allr=[];
    ChangeInt2_allr=[];
    t_pred=[];
    ChangeInt1=[];
    t_std_pred=[];
    FeatSet_Non=[];
test_data = randi(length(test_point),1,200); % esting set
    end
    for ii=1:1:1
        objId
        ii
        test_idx = test_point(test_data(ii));
        dat = dat0(:,1:test_idx);
        datSt = (Data{objId}(1:test_idx,1)==1);
        %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

        Xn = dat0(:,test_idx+1);
        XnSt = (Slp_ann(test_idx+1,1)==1);  % Apnea Status of the new
point
```

```
        dim = size(dat,1);
        dpm = dpm_gp(dat,dim,aa, s0, ss, numiter);
        pp = zeros(dpm.KK,1);
        dd = DimLVM+DimPSD;
        % cluster = cell(dpm.KK,1);
        fm = zeros(dpm.KK,1);
        fvar = zeros(dpm.KK,1);
        %%For the Prognosis part
        t_est = [];
        pp=[];
        t_std_est=[];
        test1=[];
        MaxDia =[];
        Max2Dia =[];
        Dis2mu =[];
        for j=1:dpm.KK %find the likelihood that Xn belongs to cluster
j
             [mu,sigma] = map(dpm.qq{j});  %% MAP estimation
            pp(j)=mvnpdf(Xn,mu,sigma);
            D2mu(j) = (sum(abs(Xn-mu).^2,1)).^0.5;
            ChangeAll= [];
            idx{j} =find(dpm.zz==j);
        end
    end
 sig_l = size(dat,2);
adj =zeros(sig_l,sig_l);
for iter1 = 1:1:sig_l-1
    adj(iter1, iter1+1)=1;
end
all_grp = idx; % array of all group;
grp_size = dpm.nn;

%% Use the random coordinate

for grp=1:1:length(grp_size)
n(grp) =  grp_size(grp); % number of noded
angl(grp) = 2*pi/n(grp); % rotation angle
end

angl_g = 2*pi/length(all_grp);
ord = zeros(length(all_grp),1);

for k=1:sig_l
    for id=1:1:length(grp_size)
     if ~isempty(intersect(k,all_grp{id}))
         ord(id) = ord(id)+1;
       x(k) = real(exp(angl(id)*(ord(id)-1)*i))+10*real(exp(angl_g*(id-
1)*i));
       y(k) = imag(exp(angl(id)*(ord(id)-1)*i))+10*imag(exp(angl_g*(id-
1)*i));
       continue;
     end
    end
```

```
end
    od =1:1:sig_l;
    apn_point= od(Slp_ann(1:sig_l)==1);
    non_point= od(Slp_ann(1:sig_l)==0);
    for tt=1:1:length(idx)
        apn_cl{tt}= intersect(apn_point,idx{tt})  ; % find the index of
the apnea point in the cluster tt
        non_cl{tt}= intersect(non_point,idx{tt})  ; % find the index of
the nonapnea point in the cluster tt
    end

figure,
cmap = flipud(colormap('Jet'));
for iter11=1:length(idx)

for k=apn_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);

plot(x(k),y(k),'Marker','o','MarkerFaceColor',cmap(icolor,:),'MarkerEdg
eColor',cmap(icolor,:),'MarkerSize',20)
  text(x(k),y(k),strcat('v',num2str(k)));
hold on;
end
for k=non_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);
plot(x(k),y(k),'Marker','o','MarkerEdgeColor',cmap(icolor,:),'MarkerSiz
e',20);
  text(x(k),y(k),strcat('v',num2str(k)));

%   hold off; hold on;
end
end
edges=find(adj>0);
set(gcf,'Color',[1,1,1])

for e=1:length(edges)
    [ii,jj]=ind2sub([sig_l,sig_l],edges(e));
    arrow([x(ii) y(ii)],[x(jj) y(jj)],1,'BaseAngle',60);
hold on;
end
axis off;

%% Use distance matrix coordinator

Dist1 = pdist(SP,'euclidean'); %Distance in state space
[Y11,e11] = cmdscale(Dist1);   %
x=(Y11(:,1));
y=(Y11(:,2));
z=(Y11(:,3));
     od =1:1:sig_l;
    apn_point= od(Slp_ann(1:sig_l)==1);
    non_point= od(Slp_ann(1:sig_l)==0);
```

```
    for tt=1:1:length(idx)
        apn_cl{tt}= intersect(apn_point,idx{tt})  ; % find the index of
the apnea point in the cluster tt
        non_cl{tt}= intersect(non_point,idx{tt})  ; % find the index of
the nonapnea point in the cluster tt
    end

figure,
cmap = flipud(colormap('Jet'));
for iter11=1:length(idx)
for k=apn_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);

plot3(x(k),y(k),z(k),'Marker','o','MarkerEdgeColor',cmap(icolor,:),'Mar
kerFaceColor',cmap(icolor,:),'MarkerSize',15)
  text(x(k)+0.1,y(k)+0.1,z(k)+0.1,strcat('v',num2str(k)),'FontSize',5);
hold on;
end
for k=non_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);
% %
plot3(x(k),y(k),z(k),'Marker','o','MarkerFaceColor','w','MarkerSize',15
);

text(x(k)+0.07,y(k)+0.07,z(k)+0.07,strcat('v',num2str(k)),'FontSize',5)
;
hold on;
end
end
edges=find(adj>0);
set(gcf,'Color',[1,1,1])

for e=1:length(edges)
    [ii,jj]=ind2sub([sig_l,sig_l],edges(e));
    arrow([x(ii) y(ii) z(ii)],[x(jj) y(jj) z(jj)],2.5,'BaseAngle',60);
hold on;
end
grid on;

%% Use Laplacian Matrix

Dist1 = squareform(exp(-pdist(SP,'euclidean'))); %Distance in state
space
[deg,indeg,outdeg]=degrees(Dist1);
LapMat = diag(deg)-Dist1;
[V,D] = eig(LapMat)  %
x=(V(:,2));
y=(V(:,3));
%Adjust for demostration
x(110)=0.15;
y(110)=-0.012;

x(102)= 0.16;
```

```
y(102)=-0.012

x(46)=-0.11;
y(46)=-0.02;

x(62)=-0.1;
y(62)=0.08;

y(70)=0.18;

y(78)=0.19;

x(94)=0.07;
y(94) = 0.15;

x(86) = 0.074;
y(86) = 0.16;

x(118) = 0.05;
y(118) = -0.053;
od =1:1:sig_l;
    apn_point= od(Slp_ann(1:sig_l)==1);
    non_point= od(Slp_ann(1:sig_l)==0);
    for tt=1:1:length(idx)
        apn_cl{tt}= intersect(apn_point,idx{tt})  ; % find the index of
the apnea point in the cluster tt
        non_cl{tt}= intersect(non_point,idx{tt})  ; % find the index of
the nonapnea point in the cluster tt
    end
figure,
cmap = flipud(colormap('Jet'));
for iter11=1:length(idx)
for k=non_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);
  plot(x(k),y(k),'Marker','o','MarkerSize',15);
  text(x(k)+0.007,y(k)+0.007,strcat('v',num2str(k)),'FontSize',5);
hold on;
end
for k=apn_cl{iter11}
icolor= ceil(64/length(idx)*iter11);
plot(x(k),y(k),'Marker','o','MarkerEdgeColor',cmap(icolor,:),'MarkerFac
eColor',cmap(icolor,:),'MarkerSize',15)
  text(x(k),y(k),strcat('v',num2str(k)),'FontSize',5);
hold on;
end
end
edges=find(adj>0);
set(gcf,'Color',[1,1,1])

for e=1:length(edges)
    [ii,jj]=ind2sub([sig_l,sig_l],edges(e));
    arrow([x(ii) y(ii)],[x(jj) y(jj)],3,'BaseAngle',60);
```

```
hold on;
end
axis off;
end

%% State Space trajectory only

Dist1 = pdist(SP,'euclidean'); %Distance in state space
[Y11,e11] = cmdscale(Dist1);   %
x=(Y11(:,1));
y=(Y11(:,2));
z=(Y11(:,3));
od =1:1:sig_l;
    apn_point= od(Slp_ann(1:sig_l)==1);
    non_point= od(Slp_ann(1:sig_l)==0);
    for tt=1:1:length(idx)
        apn_cl{tt}= intersect(apn_point,idx{tt})  ; % find the index of
the apnea point in the cluster tt
        non_cl{tt}= intersect(non_point,idx{tt})  ; % find the index of
the nonapnea point in the cluster tt
    end

figure,
cmap = flipud(colormap('Jet'));
for iter11=1:length(idx)

for k=apn_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);
    plot3(x(k),y(k),z(k),'Linewidth',4)

hold on;
end
for k=non_cl{iter11}
  icolor= ceil(64/length(idx)*iter11);
% %
plot3(x(k),y(k),z(k),'Linewidth',4);

hold on;
end
end
edges=find(adj>0);
set(gcf,'Color',[1,1,1])

for e=1:length(edges)
    [ii,jj]=ind2sub([sig_l,sig_l],edges(e));
    arrow([x(ii) y(ii) z(ii)],[x(jj) y(jj) z(jj)],0.1,'BaseAngle',60);
hold on;
end

grid on;
    Slp_ann(81:95)=1;
    Slp_ann(98:102)=1;
```

```
    apn_point= test_point1(Slp_ann(test_point1)==1);
    non_point= test_point1(Slp_ann(test_point1)==0);
```

A.7 Estimation of time to failure distribution

```
%% Prepare the data

load L:\Research\Hoang-Prognosis\MatlabCodes\20ApnPatients.mat; %Sleep
apnea annotation, NPSD, and LVM
load('L:\Research\Hoang-Prognosis\MatlabCodes\TimeDelay.mat') %Time
delay file, identified by using MI
combIdx = 0;
patientComb =[5,10,20]; % Index of the same patient
for longpatientID = patientComb
    combIdx = combIdx+1;
    objId = longpatientID;
    dirpath= strcat('L:\Research\Hoang-
Prognosis\MatlabCodes\DataGood_Feature1\','ID',num2str(objId),'\');
    data = Data{objId}(:,3); % Power spectrum density
    Slp_ann{combIdx} = Data{objId}(:,1);
    avg = mean(data);
    stdv = std(data);
    SleepAnno = Data{objId}(:,1);
    data = (data-avg)/stdv;
    LVM1 = Data{objId}(:,2); %Longest vertical length
    avg1 = mean(LVM1);
    stdv1 = std(LVM1);
    LVM = (LVM1-avg1)/stdv;

%% Multivariate state space reconstructions

    TauPSD = TimeDelay(longpatientID,1);
    TauLVM = TimeDelay(longpatientID,2);
    %%% tunable parameters, aa, s0, ss
    aa = 0.3;
    s0 = 1;
    ss = 0.5;
    numiter = 5;
    radius = 0.1; %radius of the neighbors 5% around the new point
    DimPSD = 10;
    DimLVM = 10;
    N= length(data);
    minimumx = min(LVM1);
    maximumx = max(LVM1);
    intervalx = maximumx-minimumx;
    LVM1=(LVM1-minimumx)/intervalx;
    avg1 = mean(LVM1);
    stdv1 = std(LVM1);
    LVM = (LVM1-avg1)/stdv;

    NX1 = N-TauPSD*(DimPSD-1);
```

```
    NX2 = N-TauLVM*(DimLVM-1);
    NX = min(NX1,NX2);
    phase1 = PhaseSpace(data,DimPSD,TauPSD);
    phase2 = PhaseSpace(LVM,DimLVM,TauLVM);
    SP{combIdx} = [phase1(1:NX,:),phase2(1:NX,:)];
    train_point = min(NX1,NX2)-30; % number of minimum training points
    test_point1= train_point-1 + ceil((NX-train_point) * rand(NX-
train_point,1));  % Select extra random testing point from 300th point
    apn_point{combIdx}= test_point1(Slp_ann{combIdx}(test_point1)==1);
    non_point{combIdx}= test_point1(Slp_ann{combIdx}(test_point1)==0);
end

%% For Non_apnea point

test_point = non_point;
for  combIdx=1:1:length(patientComb)
    if (isempty(test_point{combIdx}))
        continue;
    end
    t_pred_allr=[];
    t_std_pred_allr=[];
    ChangeInt1_allr=[];
    t_pred1_allr=[];
    t_std_pred1_allr=[];
    ChangeInt2_allr=[];
    t_pred=[];
    ChangeInt1=[];
    t_std_pred=[];
    FeatSet_Non=[];
    test_data = randi(length(test_point{combIdx}),1,20); % Testing set
    for ii=1:1:length(test_data)
        SPTemp = SP;        % Initialize state space
        test_idx = test_point{combIdx}(test_data(ii));
        Xn = SPTemp{combIdx}(test_idx+1,:);   %remove points
        SPTemp{combIdx}= SPTemp{combIdx}(1:test_idx,:); %remove points
        dat0= [SPTemp{1};SPTemp{2};SPTemp{3}];
        dat = dat0';
        XnSt = (Slp_ann{combIdx}(test_idx+1,1)==1);  % Apnea Status
        dim = size(dat,1);
        dpm = dpm_gp(dat,dim,aa, s0, ss, numiter);
        pp = zeros(dpm.KK,1);
        dd = DimLVM+DimPSD;
        fm = zeros(dpm.KK,1);
        fvar = zeros(dpm.KK,1);

%%For the Prognosis part

        t_est = [];
        pp=[];
        t_std_est=[];
        test1=[];
        MaxDia =[];
        Max2Dia =[];
```

```
        Dis2mu =[];
        % Points in cluster
         for j=1:dpm.KK
            le1 =size(SPTemp{1},1);
            le2 =size(SPTemp{2},1);
            le3 =size(SPTemp{3},1);
            idx1 =find(dpm.zz(1:le1)==j); Xn condition
            idx2 =find(dpm.zz(le1+1:le1+le2)==j);
            idx3 =find(dpm.zz(le1+le2+1:le1+le2+le3)==j);
            idx1_long{j}= intersect(find(Slp_ann{1}()==0),idx1);
            idx2_long{j}= intersect(find(Slp_ann{2}()==0),idx2);
            idx3_long{j}= intersect(find(Slp_ann{3}()==0),idx3);
      end
        clust_tran= zeros(dpm.KK,dpm.KK);
        clust_tran0 = zeros(dpm.KK,dpm.KK);
        for i= 1:1:dpm.KK
            for j=1:1:dpm.KK
               for tt=1:1:length(idx1_long{i})
                   for uu=1:1:length(idx1_long{j})
                       if (idx1_long{i}(tt)-idx1_long{j}(uu)==-1)
           clust_tran(i,j)= clust_tran(i,j)+1; % transition from i to j
           clust_tran0(i,j)= clust_tran(i,j)+1; % transition from i to
j
                       end
                       if i==j
                           clust_tran0(i,j)= 0; % transition from i to
i
                       end
                   end
               end
            end
        end
        clust_tran= clust_tran/sum(sum(clust_tran));
        clust_tran0= clust_tran0/sum(sum(clust_tran0));
        for j=1:dpm.KK %find the likelihood that Xn belongs to cluster
j
            [mu,sigma] = map(dpm.qq{j});  %% MAP estimation
            pp(j)=mvnpdf(Xn,mu',sigma);
            %     cluster{j} = find(dpm.zz==j);
            D2mu(j) = (sum(abs(Xn'-mu).^2,1)).^0.5;
            ChangeAll= [];
            le1 =size(SPTemp{1},1);
            le2 =size(SPTemp{2},1);
            le3 =size(SPTemp{3},1);
            idx1 =find(dpm.zz(1:le1)==j);% Xn condition

            idx2 =find(dpm.zz(le1+1:le1+le2)==j);
            idx3 =find(dpm.zz(le1+le2+1:le1+le2+le3)==j);
            apn_cl1= intersect(find(Slp_ann{1}()==1),idx1);
            apn_cl2= intersect(find(Slp_ann{2}()==1),idx2);
            apn_cl3= intersect(find(Slp_ann{3}()==1),idx3);
            non_cl1= intersect(find(Slp_ann{1}()==0),idx1);
            non_cl2= intersect(find(Slp_ann{2}()==0),idx2);
```

```
            non_cl3= intersect(find(Slp_ann{3}()==0),idx3);
            if length(non_cl1)== 0 ,non_cl1=[];end; % 1-by-0 error
            if length(non_cl2)== 0 ,non_cl2=[];end; % 1-by-0 error
            if length(non_cl3)== 0 ,non_cl3=[];end; % 1-by-0 error
        ChangePoint1 =[];
        for loop1=1:1:length(non_cl1)
            ChangePoint1(loop1) =
StateChange(Slp_ann{1},non_cl1(loop1));
        end

        ChangePoint2 =[];
        for loop2=1:1:length(non_cl2)
            ChangePoint2(loop2) =
StateChange(Slp_ann{2},non_cl2(loop2));
        end
       ChangePoint3 =[];
        for loop3=1:1:length(non_cl3)
            ChangePoint3(loop3) =
StateChange(Slp_ann{3},non_cl3(loop3));
        end
        TimeF{j} = [ChangePoint1,ChangePoint2,ChangePoint3];
        if (~isempty(TimeF{j}))
            [n,xout]= hist(TimeF{j},ceil(sqrt(length(TimeF{j}))));
        else
            xpdf= 0;
            ypdf= 0;
        end
        t_est(j) = xout*n'/sum(n);
        end
    pb = pp/sum(pp);
    t_pred1{combIdx}(ii) = sum(pb.*t_est);
    ChangeInt2{combIdx}(ii) = StateChange(Slp_ann{combIdx},test_idx);
    end
end
```

VITA

Trung Quoc Le

Candidate for the Degree of

Doctor of Philosophy

Thesis: A NONLINEAR STOCHASTIC DYNAMIC SYSTEMS APPROACH FOR PERSONALIZED PROGNOSTICS OF CARDIORESPIRATORY DISORDERS

Major Field: Industrial Engineering and Management

Biographical:

Personal information: Born in Da Nang City, Vietnam.

Education:

-Completed the requirements for the Doctor of Philosophy in Industrial Engineering and Management at Oklahoma State University, Stillwater, Oklahoma in July, 2013.

-Received the Bachelor of Engineering degree in Electrical-Electronics Engineering at Vietnam National University-Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam in 2006.

Experience:

-Research associate at School of Industrial Engineering and Management, Oklahoma State University from August 2008 to July 2013.

-Research intern at General Motors Research and Development from May to August 2011.

-Lecturer at Vietnam Aviation Academy from September 2006 to September 2008.

Professional Memberships:

-Institute for Operations Research and the Management Sciences (INFORMS)

-The Industrial Engineering Honor Society (Alpha Pi Mu)

-Institute of Electrical and Electronics Engineers (IEEE)

-Institute for Industrial Engineers (IIE)

-American Society of Quality (ASQ)